

A simple change detection scheme

Igor V. Nikiforov*

Laboratoire LM2S, Université de Technologie de Troyes, 12, rue Marie Curie - BP 2060 - 10010, Troyes Cedex, France

Received 12 April 1999; received in revised form 28 June 2000

Abstract

The problem of detecting changes in a stochastic system is addressed. When the model parameters after the change are unknown the generalized likelihood ratio (GLR) scheme is usually used to solve the problem. This scheme is asymptotically optimal but it is also particularly time-consuming which makes questionable its real time implementation. The *window-limited* GLR scheme, which takes into account only significant (for the detection) previous observations, is less demanding but often it is still time-consuming. In this paper we introduce an alternative approach to reduce the computational burden of the GLR scheme. The idea of this solution is to decompose a given parameter space into several subsets so chosen that in each subset the detection problem can be solved with loss of a small part of optimality by a recursive change detection algorithm. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Change detection; Suboptimal test; Unknown parameter; Fast algorithm

1. Introduction

The problem of *abrupt change detection* in systems with random disturbances has received extensive research attention in adaptive signal processing; for fault detection in technological processes; for industrial quality control; for prediction of natural catastrophic events; for monitoring in biomedicine. Let us assume that the discrete time stochastic system

$$Y_n = \mathcal{F}(X_n, \theta) + \xi_n, \quad (1)$$

where Y_n is the output signal, X_n is the measured input, θ is the model parameter and ξ_n is zero-mean

Gaussian noise, is observed sequentially, i.e. at the time n we have at our disposal n pairs of observations: $(Y_1, X_1), (Y_2, X_2), \dots, (Y_n, X_n)$. Before the unknown change time ν , the model parameter θ is equal to $\theta_0 \in \Theta \subseteq \mathbb{R}^r$, and after the change it is equal to $\theta_1 \in \Theta_1 \subset \Theta$, where $\theta_0 \notin \Theta_1$ (see Fig. 1). The domain Ω is an indifference (dead) zone which separates the point θ_0 and the subset Θ_1 . The parameter θ_0 and the domain Θ_1 are known but θ_1 and the nonrandom change time ν are unknown. Any change detection algorithm is defined by its *stopping time* N which is the time when the algorithm signals that the change has occurred. Our goal is to design the stopping time N which detects quickly any significant change in the parameter θ with low false alarm rate. Because the parameter θ_1 is unknown, the generalized likelihood ratio (GLR) scheme is usually used in this case. The first who proposed to use the GLR test as a stopping

*Tel.: + 33-3-25-71-56-78; fax: + 33-3-25-71-56-99.

E-mail address: nikiforov@univ-troyes.fr (I.V. Nikiforov).

Nomenclature

X	observations, measured input
$y(Y)$	observations, measured output
e	residual
$\ X\ = \sqrt{\sum_{i=1}^n x_i^2}$	norm of X
N	alarm (stopping) time
$\mathcal{N}(\theta, \Sigma)$	normal law with mean vector θ and covariance matrix Σ
$\mathbb{E}(Y)$	expectation of the random value Y
$\varphi(x)$	probability density
\mathbb{F}	distribution of observations
ν	change time
$\bar{\mathbb{E}}(N)$	mean time delay for detection (discrete time)
γ	mean time before a false alarm (discrete time)
$S_{n,k}(\theta_1)$	log likelihood ratio between the hypotheses $\mathcal{H}_0: \theta = \theta_0$ and $\mathcal{H}_1: \theta = \theta_1$
$\rho(\theta_1)$	Kullback–Leibler information number
d	signal-to-noise ratio
ε	coefficient of nonoptimality

Subscripts

n, k, i current time instants (discrete time)

rule was Lorden [11,12]:

$$\hat{N} = \inf \left\{ n > \tilde{m} : \max_{1 \leq k \leq n - \tilde{m}} \sup_{\theta_1 \in \Theta_1} S_{n,k}(\theta_1) \geq h \right\}, \quad (2)$$

where $S_{n,k}(\theta_1) = \sum_{j=k}^n \log(\varphi_{\theta_1}(Y_j|X_j)/\varphi_{\theta_0}(Y_j|X_j))$ is the log-likelihood ratio (LR) between the hypotheses $\mathcal{H}_0 : \theta = \theta_0$ and $\mathcal{H}_1 : \theta = \theta_1$, h is a threshold and the parameter \tilde{m} is chosen to ensure that

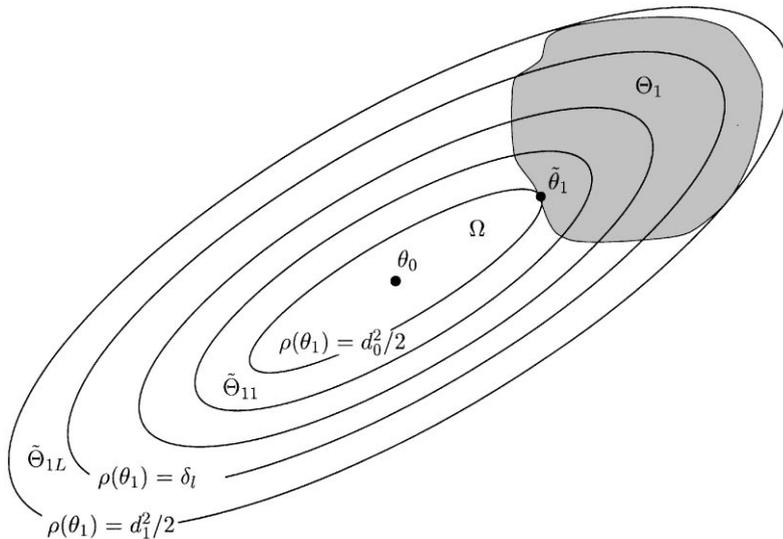


Fig. 1. The parameter space.

a maximum-likelihood estimate (MLE) of θ_1 exists. This MLE is based on $n - k + 1 \geq \tilde{m} + 1$ last observations $(Y_{n-k}, X_{n-k}), \dots, (Y_n, X_n)$. As pointed out by Lorden [11,12], Lai [8,9] and Siegmund and Venkatraman [25], this scheme is *asymptotically optimal*, therefore it minimizes the mean time delay for detection simultaneously for all $\theta_1 \in \Theta_1$ over the class of stopping rules satisfying a given false alarm rate. Nevertheless, the GLR scheme has an obvious disadvantage. At every time n this scheme involves n maximizations of the LR over $\theta_1 \in \Theta_1$. It is easy to see that the number of maximizations at time n grows to infinity with n . If we assume a given mean time before a false alarm γ as a typical period of observation then the mean number of maximization of the LR over $\theta_1 \in \Theta_1$ that should be performed at every time $n \in [1, \gamma]$ is $\gamma/2$.

Several methods have been proposed to reduce the computational cost of the GLR scheme (recent results are summarized by Basseville and Nikiforov [4] and Lai [8,9]). Willsky and Jones [28] introduced the *window-limited* GLR (WL GLR) scheme:

$$\hat{N}_{m, \tilde{m}} = \inf \left\{ n > \tilde{m} : \max_{\max\{0, n-m\} + 1 \leq k \leq n - \tilde{m}} \sup_{\theta_1 \in \Theta_1} S_{n,k}(\theta_1) \geq h \right\}, \quad (3)$$

$$0 \leq \tilde{m} < m,$$

which was theoretically investigated and justified by Lai [8,9]. The idea of Willsky–Jones–Lai is to reduce the computational burden in scheme (2) by carrying out only the m last observations $(Y_{n-m+1}, X_{n-m+1}), \dots, (Y_n, X_n)$ and rejecting the observations $(Y_1, X_1), (Y_2, X_2), \dots, (Y_{n-m}, X_{n-m})$ which are not important for the detection. It is known that the Kullback–Leibler information number $\rho(\theta_1, \theta_0)$ between two probability densities φ_{θ_1} and φ_{θ_0} plays a crucially important role in sequential detection (see details in [4]). Because θ_0 is a known and constant parameter, we will omit θ_0 in the rest of the paper. In our case the Kullback–Leibler information number $\rho(\theta_1)$ is defined as follows (often this definition of the information number $\rho(\theta_1)$ can be simplified):

$$\mathbb{E}_{\theta_1}^k \left(\frac{1}{n} S_{k+n-1,k}(\theta_1) \right) \rightarrow \rho(\theta_1) \quad (4)$$

as $n \rightarrow \infty$ uniformly in $k \geq 1$

and $\mathbb{E}_{\theta_1}^v(\cdot)$ is the expectation associated with the probability measure which ‘drives’ the model (1) after the change time v . Let us also define the minimum Kullback–Leibler information number $\rho_{\min} = \inf_{\theta_1 \in \Theta_1} \rho(\theta_1)$ which characterizes the ‘statistical distance’ between the domain Θ_1 and θ_0 . In other words, ρ_{\min} is the ‘distance’ between the ‘closest’ alternative $\tilde{\theta}_1 = \operatorname{arginf}_{\theta_1 \in \Theta_1} \rho(\theta_1)$ (see Fig. 1) and θ_0 . It follows from [8,9] that this WL GLR scheme holds asymptotic optimality if the size of the moving window is $O(\log \gamma / \rho_{\min})$ as $\gamma \rightarrow \infty$. The statistical interpretation of this result is very simple: the optimum mean time delay to detect the closest alternative $\tilde{\theta}_1$ is asymptotically equal to $\log \gamma / \rho_{\min}$, hence, to extract almost all useful information from the observations it is necessary to carry out at least $O(\log \gamma / \rho_{\min})$ last observations. Because \tilde{m} is usually small (see Section 3), the WL GLR scheme involves $O(\log \gamma / \rho_{\min})$ LR maximizations at every time n . In some situations this fact considerably reduces the computational burden (and also memory requirements) of the GLR scheme and makes this detection scheme manageable in real time implementations. Nevertheless, the WL GLR scheme is still time-consuming if the minimum Kullback–Leibler information number ρ_{\min} is small and/or the mean time before a false alarm γ is large.

The GLR scheme is equivalent to a collection of *infinite number* of the parallel CUSUM tests each of them is designed to detect a particular value θ_1 . The idea of our approach is to thin out this collection of the CUSUM tests such that this new (finite) collection of recursive tests will be almost optimal with respect to the quickest detection criterion. We propose to decompose a given parameter space into several subsets so chosen that in each subset the detection problem can be solved with loss of a small part, ε , of optimality by a recursive change detection algorithm. A collection of such parallel recursive algorithms establishes the ε -optimal detection scheme which reduces the computational cost of the GLR scheme. By choosing an acceptable value ε of *non optimality*, the designer can easily find a trade-off between the complexity of this ε -optimal change detection algorithm and its efficiency. The remaining part of this paper is organized as follows. The models and the criteria are presented

in Section 2. The outlines of our approach are discussed in Section 3. We design the ε -optimal detection algorithm in Section 4. In Section 5, we compare the ε -optimal detection scheme with the WL GLR one. Finally, some conclusions are given in Section 6.

2. Models and criteria

2.1. Models

We limit our discussion by two particular cases of the discrete time system given by Eq. (1).

2.1.1. (1) $\mathcal{F}(X_n, \theta) = \theta$

It is necessary to detect quickly any significant change at an unknown time ν in the mean vector θ of an independent Gaussian multivariate ($r > 1$) sequence

$$\mathcal{L}(Y_n) = \begin{cases} \mathcal{N}(\theta_0, \Sigma) & \text{if } n < \nu, \\ \mathcal{N}(\theta_1, \Sigma) & \text{if } n \geq \nu, \end{cases} \quad (5)$$

taken from some process. The domain Θ_1 is given by the following equation:

$$\begin{aligned} \Theta_1 &= \{\theta_1 : d_0^2 \leq d^2 \\ &= (\theta_1 - \theta_0)^T \Sigma^{-1} (\theta_1 - \theta_0) \leq d_1^2\}, \end{aligned} \quad (6)$$

where $0 < d_0 < d_1 < \infty$ are given bounds for the signal-to-noise ratio (SNR) d .¹ In the case of model (5) the Kullback–Leibler information number is $\rho(\theta_1) = \frac{1}{2}(\theta_1 - \theta_0)^T \Sigma^{-1} (\theta_1 - \theta_0)$ [4]. Hence, the domain Θ_1 limits the part of the parameter space where the Kullback–Leibler information number $\rho(\theta_1)$ varies between $d_0^2/2$ and $d_1^2/2$ and definition (6) can be expressed in term of $\rho(\theta_1)$

$$\Theta_1 = \{\theta_1 : d_0^2/2 \leq \rho(\theta_1) \leq d_1^2/2\}. \quad (7)$$

¹This is a general definition of the domain Θ_1 . Alternatively, the domain Θ_1 can be defined as a subset of the domain $\Theta : d_0^2 \leq (\theta_1 - \theta_0)^T \Sigma^{-1} (\theta_1 - \theta_0) \leq d_1^2$ such that: (i) $d_0^2/2 = \inf_{\theta_1 \in \Theta} \rho(\theta_1)$; (ii) $d_1^2/2 = \sup_{\theta_1 \in \Theta} \rho(\theta_1)$ as it is shown in Fig. 1. In this case, the rest $\Theta \setminus \Theta_1$ is considered as a part of the indifference zone Ω .

The value d_0 can be extracted from technical norms, standards, etc. and d_1 can be so chosen that the mean time delay for detection is close to one (at least one observation should be done in any case!).

2.1.2. (2) $\mathcal{F}(X_n, \theta) = X_n^T \theta$

It is necessary to detect quickly any significant change at an unknown time ν in the parameter vector θ of the regression model

$$y_n = \begin{cases} X_n^T \theta_0 + \xi_n & \text{if } n < \nu, \\ X_n^T \theta_1 + \xi_n & \text{if } n \geq \nu, \end{cases} \quad (8)$$

where y_n is the scalar output, ξ_n is an independent scalar Gaussian sequence, $\mathcal{L}(\xi_n) \sim \mathcal{N}(0, 1)$, X_n is the measured input, $X \in \mathbb{R}^r$. For the regression model (8) the domain Θ_1 is given by the following equation:

$$\Theta_1 = \{\theta_1 : d_0^2 \leq d^2 = (\theta_1 - \theta_0)^T R (\theta_1 - \theta_0) \leq d_1^2\}, \quad (9)$$

where $R \in \mathbb{R}^{r \times r}$ is a positive-definite matrix, if X_n is a stationary random sequence (X_n and ξ_n are independent) then R is a covariance matrix, given by $R \triangleq \mathbb{E}(X_n X_n^T)$, if X_n is a deterministic sequence then we assume that [30]

$$\begin{aligned} \frac{1}{n} \sum_{i=k}^{k+n-1} X_i X_i^T &\rightarrow R \\ \text{as } n \rightarrow \infty &\text{ uniformly in } k \geq 1. \end{aligned} \quad (10)$$

Let us compute the Kullback–Leibler information number $\rho(\theta_1)$ by using Eq. (4). We assume that θ_1 is the true value of θ , Eq. (10) is satisfied and compute the expectation of the log LR $1/n S_{k+n-1, k}(\theta_1)$:

$$\begin{aligned} \mathbb{E}_{\theta_1}^k \left(\frac{1}{n} S_{k+n-1, k}(\theta_1) \right) \\ = \frac{1}{2n} (\theta_1 - \theta_0)^T \sum_{i=k}^{k+n-1} X_i X_i^T (\theta_1 - \theta_0). \end{aligned} \quad (11)$$

By putting Eqs. (10) and (11) together and by taking the limit of the result as $n \rightarrow \infty$, we get the Kullback–Leibler information number

$\rho(\theta_1) = \frac{1}{2}(\theta_1 - \theta_0)^T R(\theta_1 - \theta_0)$. Therefore, analogously to the previous case, the domain Θ_1 (9) can be expressed in term of $\rho(\theta_1)$ by (7).

2.1.3. Remark

The model $\mathcal{F}(X_n, \theta) = \theta$ plays an important role in the theory of change detection because frequently the detection of *additive* changes in state-space, regression, and ARMA models, subspace-based detection can be reduced (sometimes asymptotically) to the above *basic* model by using a residual generation mechanism (see examples in Section 2.2 and more extensive discussions in [4,8,3]). The model $\mathcal{F}(X_n, \theta) = X_n^T \theta$ is widely used in control applications and also in statistical signal processing (trajectography, adaptive arrays, adaptive filters, etc.). In the second case the signal-processing algorithms often carry out narrow-band signals. The complex field is the natural domain for handling such signals which contain information in the carrier phase and the amplitude. Hence, the scalar $y \in \mathbb{C}$ and the vectors $X, \theta \in \mathbb{C}^r$ are complex in this case. In the present paper the case of complex signals will not be developed, we note only that the extension of the proposed change detection scheme to the case of complex signal can be easily done by using the technique of a complex gradient operator [7].

2.2. Examples and motivations

The goal of this subsection is to briefly discuss a practical application: the navigation systems integrity monitoring. This discussion gives a physical insight into the problem under consideration and introduces the formal definition of criteria and the concept of ε -optimality given in Section 2.3.

2.2.1. Navigation systems integrity monitoring

For many safety-critical applications, a major problem of the existing navigation systems consists in its lack of integrity. Integrity monitoring concept defined by the International Civil Aviation Organization (ICAO) requires that a navigation system detects faults and removes them from the navigation solution before they sufficiently contaminate the output. We use here the model $\mathcal{F}(X_n, \theta) = \theta$.

For this reason our discussion is limited by two types of navigation systems: the strapdown inertial reference unit (SIRU) and the global positioning system (GPS).

2.2.2. The model of SIRU with a degradation

Conventional SIRU incorporates $s \geq 4$ single degree-of-freedom sensors (laser giros or accelerometers) [26]. We assume that n skewed axis inertial sensors are equally spaced on a cone with half-angle $\alpha = 54.736^\circ$. In this case the measurement model of SIRU is defined by the following static regression model with redundancy [13]:

$$Y_n = H \mathcal{A}_n + \xi_n + Y(n, v), \quad (12)$$

where $\mathcal{A}_n \in \mathbb{R}^3$ is a nonrandom unknown state vector (say, acceleration), $Y_n \in \mathbb{R}^s$ is a vector of measurements, $\xi_n \in \mathbb{R}^s$ is a Gaussian white noise with zero mean and covariance $R = \sigma^2 I_s$, $\sigma^2 > 0$, $H = (h_{ij})$ is a matrix of size $s \times 3$, $h_{i1} = \cos \beta_i$, $h_{i2} = \sin \beta_i \sin \alpha$, $h_{i3} = -\cos \alpha$, $\beta_i = (360^\circ/s)(i - 1)$, $i = 1, \dots, s$ and $Y(n, v)$ is a fault occurring at time v , namely:

$$Y(n, v) = \begin{cases} 0 & \text{if } n < v, \\ Y & \text{if } n \geq v. \end{cases} \quad (13)$$

2.2.3. The model of GPS with a degradation

The GPS navigation solution is based upon accurate measuring the distance (*range*) from s visible satellites with known locations $X_i = (x_i, y_i, z_i)^T$, $i = 1, \dots, s$, to a user (vehicle) at $X_u = (x_u, y_u, z_u)^T$. The distance from the i th satellite to the user is defined as $d_i = \|X_i - X_u\|$. The *pseudo-range* (i.e. measure of the distance) r_i from the i th satellite to the user can be written as $r_i = d_i + cb + \xi_i$, $i = 1, \dots, s$, where $b \in \mathbb{R}$ is a user clock bias, $c \simeq 2.9979 \times 10^8$ m/s is the speed of light and ξ_i is an additive pseudo-range error at the user's position. Let us introduce the following vectors: $R = (r_1, \dots, r_s)^T$ and $X = (X_u^T, b)^T$. By linearizing the pseudo-range equation with respect to the state vector X around the working point X_0 , we get the measurement equation

$$Y = R - R_0 \simeq Hx + \xi, \quad x = X - X_0,$$

where $R_0 = (r_{1_0}, \dots, r_{s_0})^T$, $r_{i_0} = \|X_i - X_{u_0}\| + cb_0$, $\xi = (\xi_1, \dots, \xi_n)^T$, $H = \partial R / \partial X|_{X=X_0}$ is the Jacobian matrix of size $s \times 4$ and $\xi \in \mathbb{R}^s$ is a Gaussian white noise with zero mean and covariance $R = \sigma^2 I_s$, $\sigma^2 > 0$. The degradation of GPS channels is represented by an additional biases in the pseudo-ranges [18]:

$$Y_n = Hx_n + \xi_n + Y(n, \nu). \quad (14)$$

Let us assume that at least five satellites are visible ($s \geq 5$). As it follows from Eqs. (12) and (14), the optimal estimate of the vehicle's acceleration \mathcal{A}_n or the user's fix x is given by the least-squares (LS) algorithm

$$\hat{\mathcal{A}}_n = (H^T H)^{-1} H^T Y_n \quad \text{or} \quad \hat{x}_n = (H^T H)^{-1} H^T Y_n. \quad (15)$$

As it follows from Eqs. (12), (14) and (15), a fault Y , affecting the sensors (or channels), implies an additional error $\mathbb{E}(\hat{\mathcal{A}}_n - \mathcal{A}_n) = (H^T H)^{-1} H^T Y$ in the vector $\hat{\mathcal{A}}_n$ (or \hat{x}_n) which contaminates the output of the navigation system [18]. The key tool for detecting additive changes $Y(n, \nu)$ in the static regression models with redundancy (12) and (14) consists of the *transformation* of the measurement vector Y_n into the residual (or parity) vector of the analytical redundancy approach (see details in [4,13,18]).

Let us consider the model of SIRU given by Eq. (12). The transformation is given by [13]

$$e_n = T^T Y_n = T^T (H \mathcal{A}_n + \xi_n + Y(n, \nu)), \quad (16)$$

where $T^T H = 0$, $T^T T = I_{s-3}$, $T = (t_1, \dots, t_{s-3})$ is a matrix of size $s \times (s-3)$, and t_1, \dots, t_{s-3} are the eigenvectors of the projection matrix $\Pi = I - H(H^T H)^{-1} H^T$ corresponding to unit eigenvalues of the matrix Π . In fact, the parity vector e_n is the transformation of the measurements Y_n into a set of $s-3$ linearly independent variables by projection onto the left null space of the matrix H :

$$\mathcal{L}(e_n) = \begin{cases} N(0, \sigma^2 I_{s-3}) & \text{if } n < \nu, \\ N(T^T Y, \sigma^2 I_{s-3}) & \text{if } n \geq \nu. \end{cases} \quad (17)$$

Therefore, the change detection problems for the above models (12) and (14) are reduced to the change detection problem for the model $\mathcal{F}(X_n, \theta) = \theta$ given by (5).

2.2.4. Intuitive formulation of the detection criteria

Let us deduce the desirable detection criteria from the above examples. First of all, the change time ν , when the fault arrives, and the new value of the parameter vector θ_1 are not simply unknown but can be intentionally chosen to maximize their negative impacts on the considered system. The requirements of ICAO usually demand the min-max fault detection and removing with respect to its onset time ν and the new value of θ_1 . Second, it is intuitively obvious that the criterion which must be used should favor fast detection with few false alarms. In other words the detection delay $N - \nu + 1$ (which is measured by the number of faulty observations taken before the alarm has been declared), given that the fault has been detected after its onset time ν , should be stochastically small and the time before a false alarm should be stochastically large. We denote the conditional detection delay by $N - \nu + 1 | N \geq \nu$, where the condition $N \geq \nu$ means that the fault has been declared after its onset time, and the time before a false alarm by N given that $\nu = \infty$, where the event $\nu = \infty$ means that the navigation system works good. Fast detection is necessary because, between the fault onset time ν and the alarm time N , abnormal measurements are taken in the navigation systems, which is clearly very undesirable. On the other hand, false alarms result in lower accuracy of the estimates because some correct information is not used. The optimal solution involves a tradeoff between these two contradictory requirements.

2.3. Formal definition of the detection criteria

We assume the nonBayesian approach, i.e. the change time ν is an unknown but *nonrandom* integer value. The statistical performance of non-Bayesian change detection algorithms is usually measured with the aid of two criteria.

2.3.1. ARL function.

The first criterion is based on the *average run length* (ARL) function [4]:

$$\text{ARL}(\theta) \triangleq \mathbb{E}_\theta^1(N), \quad (18)$$

where $\mathbb{E}_\theta^1(N)$ is the average number of observations required for the algorithm to signal that θ has changed given the change time is $\nu = 1$. In other words the ARL function provides us with a criterion for assessing the performance of the change detection scheme under some standard conditions. The main feature of quadratic algorithms is that the ARL function depends only on the SNR d (or, on the Kullback–Leibler information number $\rho(\theta_1)$). A typical ARL surface $\text{ARL}(\theta)$, where $\theta = (\vartheta_1, \vartheta_2)$, for the GLR scheme and corresponding level lines $\text{ARL}(\vartheta_1, \vartheta_2) = \text{const}$ are shown in Fig. 2. This figure shows that the quadratic change detection scheme (GLR or CUSUM) is able to detect *any* changes $\theta_1 : (\theta_1 - \theta_0)^T \Sigma^{-1} (\theta_1 - \theta_0) = d^2$ from the nominal value θ_0 with the same mean time delay for detection $\text{ARL}(d > 0)$. The level lines of $\text{ARL}(\vartheta_1, \vartheta_2)$ are the ellipses $(\theta_1 - \theta_0)^T \Sigma^{-1} (\theta_1 - \theta_0) = d^2$ (see Fig. 2, right). Naturally, it is desirable to have an algorithm with the minimum mean time delay for detection $\text{ARL}(d > 0)$ under a given mean time before a false alarm $\text{ARL}(0)$. A drawback of this criterion is the assumption $\nu = 1$, hence, as it follows from Section 2.2, this criterion plays an auxiliary role in this paper.

2.3.2. The min–max criterion

We look for a change detection scheme (stopping time N) which is ‘independent’ of ν . Let $\mathbb{F}^\nu \nu \geq 1$

be the distribution of the observations $(Y_1, X_1), \dots, (Y_{\nu-1}, X_{\nu-1}), (Y_\nu, X_\nu), \dots$ when (Y_ν, X_ν) is the first pair of observations distributed according to the measure which ‘drives’ the stochastic system after the change. The associated expectation is denoted by \mathbb{E}^ν . The notation \mathbb{F}^∞ corresponds to the case when all observations are distributed according to the measure which ‘drives’ the stochastic system before the change ($\nu = \infty$). It is necessary to minimize the ‘worst case’ (with respect to ν) mean time delay for detection

$$\bar{\mathbb{E}}(N; d) \triangleq \sup_{\nu \geq 1} \mathbb{E}_{\theta_1}^\nu : 2\rho(\theta_1) = d^2 (N - \nu + 1 | N \geq \nu) \quad (19)$$

for a given mean time before a false alarm

$$\gamma \triangleq \mathbb{E}^\infty(N). \quad (20)$$

This criterion was introduced by Pollak and Siegmund [23] and studied by Pollak [22], Lai [8] and Yakir [29]. To simplify our presentation, we omit the words ‘worst case’ when we discuss the mean time delay for detection $\bar{\mathbb{E}}(N; d)$ in the rest of the paper.

2.3.3. The concept of ϵ -optimality

Let us pursue our discussion of the ϵ -optimality concept. First of all, we introduce the lower bound of the mean time delay for detection. Because the

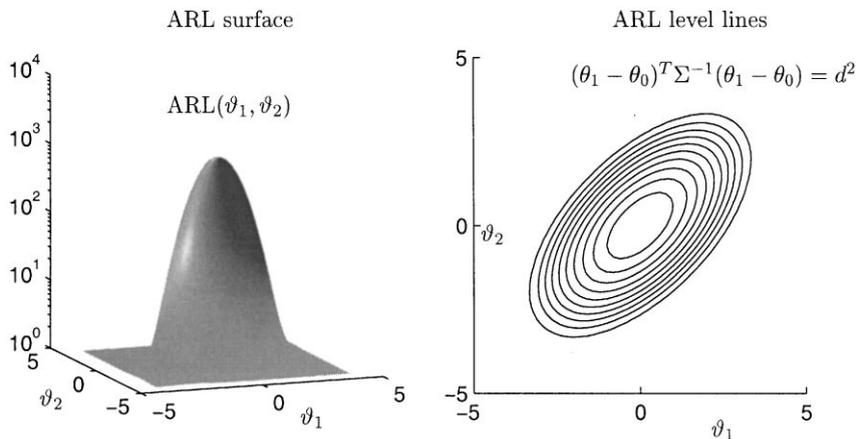


Fig. 2. Typical ARL surface and level lines for the quadratic change detection algorithm.

fast detection and few false alarms are contradictory requirements, the traditional solution is to bound the mean time before a false alarm $\mathbb{E}^\infty(N)$ from below by a positive constant γ and to define the smallest possible mean time delay for detection $n(\gamma)$ under this constraint. Therefore, we define the class \mathcal{K}_γ of all tests N for which the mean time before a false alarm is bounded from below by γ , namely: $\mathcal{K}_\gamma = \{N : \mathbb{E}^\infty(N) \geq \gamma\}$. We next define the lower bound $n(\gamma)$ of the mean time delay for detection $\bar{\mathbb{E}}(N)$ as the infimum of $\bar{\mathbb{E}}(N)$ in the class \mathcal{K}_γ : $n(\gamma) \triangleq \inf\{\bar{\mathbb{E}}(N) : N \in \mathcal{K}_\gamma\}$. Because the SNR d and the mean time before a false alarm γ can vary, it is convenient for us to consider the mean time delay for detection $\bar{\mathbb{E}}(N)$ and its lower bound n as functions of d and γ :

$$(d, \gamma) \mapsto \bar{\mathbb{E}}(N; d, \gamma) \quad \text{and} \quad (d, \gamma) \mapsto n(d, \gamma).$$

The algorithm which attains the lower bound is called optimal. If an algorithm attains the lower bound for all values of d , then it is called uniformly optimal. It can happen that the uniformly optimal algorithm does not exist or it is very time-consuming (as the GLR scheme), in spite of this fact, the lower bound plays a crucially important role in the

ε -optimality concept because it represents a kind of invariant to adequately appreciate the quality of a change detection algorithm. As it follows from [8], this lower bound is given by

$$n(d, \gamma) \sim \frac{2 \log \gamma}{d^2} \quad \text{as } \gamma \rightarrow \infty, \tag{21}$$

where $d^2 = 2\rho(\theta_1)$. To explain the above asymptotic formula, let us recall the definition of the symbol \sim . We consider two functions: $f, g : \mathbb{R}_+ \rightarrow \mathbb{R}$, $f(x) \sim g(x)$ as $x \rightarrow \infty$ means that $f(x) = \lambda(x)g(x)$, where $\lambda(x) \rightarrow 1$ as $x \rightarrow \infty$.

The ε -optimality concept is shown in Fig. 3. Let us assume that N is a test from the class \mathcal{K}_{γ_0} , i.e. $\mathbb{E}^\infty(N) \geq \gamma_0$. Let us define the loss of optimality $\varepsilon(N; d, \gamma) = 1 - (n(d, \gamma) / \bar{\mathbb{E}}(N; d, \gamma))$ of N . For given values of d and γ , the quantity $\varepsilon(N; d, \gamma)$ shows the relative efficiency of N with respect to the lower bound $n(d, \gamma)$. The closer the coefficient ε to zero, the better is the test N . If $\varepsilon = 0$, then the test N is optimal. Typical graphs of the functions $d \mapsto n(d, \gamma_0)$ and $d \mapsto \bar{\mathbb{E}}(N; d, \gamma_0)$ are shown in Fig. 3, where the function $d \mapsto n(d, \gamma_0)$ is shown by a solid line and the function $d \mapsto \bar{\mathbb{E}}(N; d, \gamma_0)$ is shown by a dash-dot line. It is easy to get an optimal test N for a given

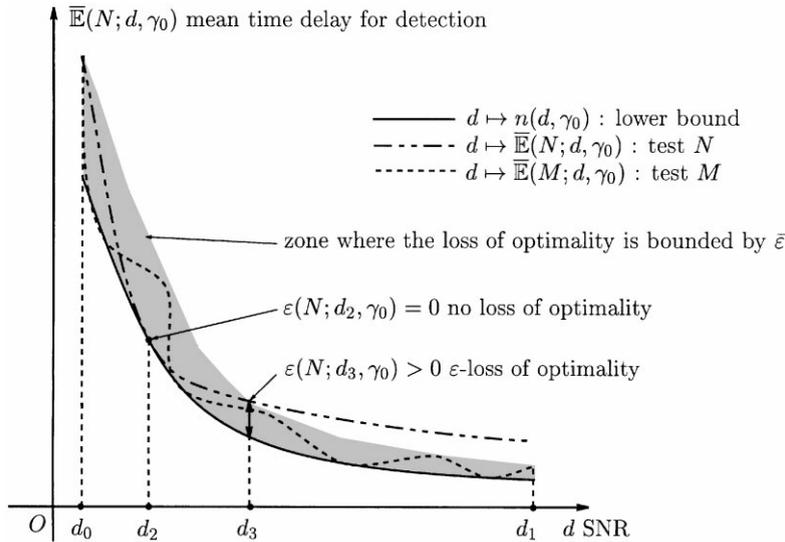


Fig. 3. The concept of ε -optimality.

value of the SNR, say d_2 (see Fig. 3), hence $\varepsilon(N; d_2, \gamma_0) = 0$, i.e. no loss of optimality. But for another value of the SNR, say $d_3 \neq d_2$, the test N is not optimal and this leads to $\varepsilon(N; d_3, \gamma_0) > 0$, i.e. ε -loss of optimality (see Fig. 3).

Let us fix the maximum acceptable loss of optimality (coefficient of nonoptimality) $\bar{\varepsilon}$ and define the following function $f: d \mapsto f(d, \gamma) = n(d, \gamma)/(1 - \bar{\varepsilon})$. It is easy to see that the region between the graphs of $d \mapsto n(d, \gamma_0)$ and $d \mapsto f(d, \gamma_0)$ represents the zone where the loss of optimality is bounded by $\bar{\varepsilon}$: $\varepsilon(N; d, \gamma_0) \leq \bar{\varepsilon}$. This region is shown by a shaded area in Fig. 3. To characterize the loss of optimality $\varepsilon(N)$ for the test N as d ranges over the interval $[d_0; d_1]$ we propose the following. Because all the theoretical results on the optimal detection have an asymptotic sense, we, first, compute the limit $\varepsilon(N; d)$ of the function $\varepsilon(N; d, \gamma)$ as $\gamma \rightarrow \infty$ for a given d and, next, compute the supremum of $\varepsilon(N; d)$ as d ranges over $[d_0; d_1]$. Hence, we get

$$\varepsilon(N) \triangleq \sup_{d \in [d_0; d_1]} \varepsilon(N; d);$$

$$\varepsilon(N; d) \triangleq \lim_{\gamma \rightarrow \infty} \left(1 - \frac{n(d, \gamma)}{\bar{\varepsilon}(N; d, \gamma)} \right). \quad (22)$$

Let $0 < \bar{\varepsilon} < 1$ be a given constant (coefficient of nonoptimality). We say that the test N is $\bar{\varepsilon}$ -optimal if $\varepsilon(N) = \bar{\varepsilon}$. The definition of $\bar{\varepsilon}$ -optimality may be illustrated geometrically as in Fig. 3. Let us analyze two change detection tests: N and M which are shown by dash-dot and dotted lines, respectively. Our goal is to check their $\bar{\varepsilon}$ -optimality. The shaded area consists of all points for which the loss of optimality is bounded by $\bar{\varepsilon}$. The definition of $\bar{\varepsilon}$ -optimality asserts that the entire graph of $d \mapsto \bar{\varepsilon}(N; d, \gamma_0)$ (or the entire graph of $d \mapsto \bar{\varepsilon}(M; d, \gamma_0)$) above the interval $[d_0; d_1]$ lies within this area. As it follows from Fig. 3, the definition is satisfied for the test M , i.e. the test M is $\bar{\varepsilon}$ -optimal, and it is not satisfied for N (see Fig. 3).

3. Outlines of the proposed approach

The Kullback–Leibler information number $\rho(\theta_1)$ will be used to solve the problem of ε -optimal

change detection in the rest of the paper. To reduce the computational cost of the GLR scheme, the following plan is proposed:

- (1) The first step is to cover a given domain Θ_1 by a collection of L subsets (zones of responsibility) $\tilde{\Theta}_{11}, \dots, \tilde{\Theta}_{1L}$ as shown in Figs. 1 and 4. These subsets are defined as follows:

$$\tilde{\Theta}_{1l} = \{\theta_1 : \delta_{l-1} \leq \rho(\theta_1) \leq \delta_l\},$$

$$l = 1, \dots, L, \quad (23)$$

where $\delta_0 = \inf_{\theta_1 \in \Theta_1} \rho(\theta_1) = d_0^2/2$ and $\delta_L = \sup_{\theta_1 \in \Theta_1} \rho(\theta_1) = d_1^2/2$. Let us define now a subdivision $\sigma = \{\zeta_1, \dots, \zeta_L\}$ of the closed interval $[d_0^2/2; d_1^2/2]$, where the values ζ_l are so chosen that $\delta_0 = d_0^2/2 < \zeta_1 < \delta_1, \dots, \delta_{L-1} < \zeta_L < \delta_L = d_1^2/2$. The idea of our approach is to design L parallel recursive tests, each of them is asymptotically optimal for $\theta_1 \in S_1(\zeta_l)$, where $S_1(\zeta_l)$ is a surface defined as a set of points satisfying the equation $\rho(\theta_1) = \zeta_l$, $l = 1, \dots, L$ (see Figs. 1 and 4).

- (2) The next step is to design the asymptotically optimal recursive tests for detection of changes from θ_0 to $\theta_1 : \rho(\theta_1) = \zeta_l$. Moreover, these recursive schemes should hold their mean time delays for detection stably (with ε variations) against small changes in the Kullback–Leibler

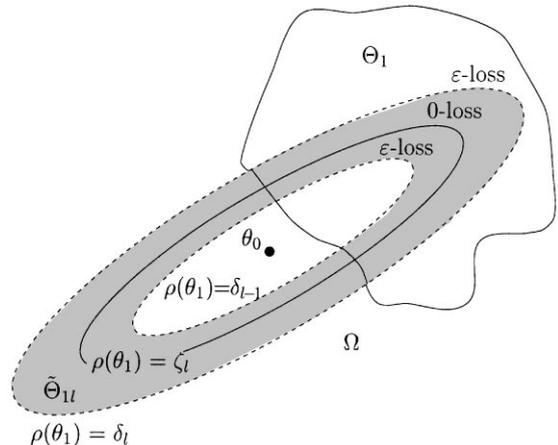


Fig. 4. The zone of responsibility.

information number. Therefore, if L zones of responsibility (23) and L tuning values ζ_1, \dots, ζ_L are so chosen that the variation of the actual value ζ_l of the Kullback–Leibler information number $\rho(\theta_1)$, where $\theta_1 \in \tilde{\Theta}_{1l}$, around the tuning value ζ_l is limited for each zone then we can expect that the collection of such recursive tests will be ε -optimal.

To simplify the notation, from now on, we consider an l th zone of responsibility and we omit the index l . To design an asymptotically optimal test, two different approaches can be used [4,27]: the GLR test and the *weighted LR (WLR)* test.

3.1. GLR test

This solution consists in maximizing the log LR $S_{n,k}(\theta_1)$ when θ_1 is restricted to lie in the subset $S_1(\zeta)$:

$$\hat{S}_{n,k} = \sup_{\theta_1: \rho(\theta_1)=\zeta} S_{n,k}(\theta_1), \quad (24)$$

which results in the constrained GLR test. A direct approach to this problem is the method of Lagrange's multipliers

$$\frac{\partial S_{n,k}(\theta_1)}{\partial \theta_1} + \lambda \frac{\partial \rho(\theta_1)}{\partial \theta_1} = 0, \quad \rho(\theta_1) = \zeta. \quad (25)$$

The above system of equations is to be solved (if possible) for $r+1$ unknowns $\vartheta_{1,1}, \dots, \vartheta_{1,r}$ and λ . Let us pursue our discussion of the parameter \tilde{m} . The log LR $S_{n,k}(\theta_1)$ in (25) is based on $n-k+1 \geq \tilde{m}+1$ last observations. Hence, the choice of the parameter \tilde{m} defines the minimum number of samples to compute the log LR $S_{n,k}(\theta_1)$. The parameter \tilde{m} follows from the statistical structure of the model considered in (1) [6]. For the first model $Y_n = \theta + \xi_n$, $\xi_n \sim \mathcal{N}(0, \Sigma)$, the parameter is $\tilde{m} = 0$, hence, the smallest number of samples is equal to 1 and this log LR $S_{n,n}(\theta_1)$ is based on the unique observation Y_n . For the second model $y_n = X_n^T \theta + \xi_n$, $\xi_n \sim \mathcal{N}(0, 1)$, the parameter is $\tilde{m} = r$, hence, the smallest number of samples is equal to $r+1$ and this log LR $S_{n,n-r}(\theta_1)$ is based on the observations $(y_{n-r}, X_{n-r}), \dots, (y_n, X_n)$.

3.2. WLR test

The second solution is to use the *weighted LR (WLR)*

$$\tilde{S}_{n,k} = \int \dots \int_{\theta_1: \rho(\theta_1)=\zeta} e^{S_{n,k}(\theta_1)} f(\theta_1) dS, \quad (26)$$

where dS is the element of the surface $S_1(\zeta)$ and $f(\theta_1) \geq 0$ may be interpreted as the *weighting function*: $\int \dots \int_{\theta_1: \rho(\theta_1)=\zeta} f(\theta_1) dS = 1$. The main difficulty here is to choose the weighting function $f(\theta_1)$. Often this choice can be done by using the invariant properties of a given family of distributions.

3.3. Recursive scheme

The decision functions $\hat{S}_{n,k}$ and $\tilde{S}_{n,k}$ have negative drifts before the change and asymptotically the same positive drifts ζ after the change $1/(n-k+1)\hat{S}_{n,k}$ and $1/(n-k+1)\tilde{S}_{n,k}$ converge with probability 1 to ζ as $n-k \rightarrow \infty$ and $k \geq v$ under the distribution which 'drives' the model (1) after the change (see [4,15,18,20] for details). Because of this asymptotic property of the decision functions $\hat{S}_{n,k}$ and $\tilde{S}_{n,k}$, a recursive decision rule can be designed for each subset $S_1(\zeta)$. It is known [4] that the CUSUM-type change detection algorithm can be represented as a *repeated* sequential probability ratio test (SPRT) in the case of two simple hypothesis ($\mathcal{H}_0: \theta = \theta_0$ vs $\mathcal{H}_1: \theta = \theta_1$). Let us consider this *repeated* SPRT (N_j, D_j):

$$N_j = \inf\{n > N_{j-1} : (S_{n, N_{j-1}+1}(\theta_1) \geq h) \cup (S_{n, N_{j-1}+1}(\theta_1) \leq 0)\}, \quad (27)$$

$$D_j = \begin{cases} 1 & \text{if } S_{N_j, N_{j-1}+1}(\theta_1) \geq h, \\ 0 & \text{if } S_{N_j, N_{j-1}+1}(\theta_1) \leq 0, \end{cases}$$

where N_j is the exit time, D_j is the decision rule (if $D_j = i$ then \mathcal{H}_i is accepted, $i = 0, 1$) and h is an upper threshold such that $0 < h < \infty$. We define the following sequence of exit times $N_0 = 0 < N_1 < \dots < N_j < \dots$. Hence, the observation is stopped after the first sample of size $N_j - N_{j-1}$ for

which the decision rule D_j is 1 (\mathcal{H}_1 is accepted). The stopping time $N_c = N_j$ of this recursive CUSUM scheme (repeated SPRT) (27) can be rewritten as

$$N_c = \inf\{n \geq 1 : S_n \geq h\},$$

$$S_n = \left(S_{n-1} + \log \frac{\varphi_{\theta_1}(Y_n|X_n)}{\varphi_{\theta_0}(Y_n|X_n)} \right)^+, \quad (28)$$

where $S_0 = 0$, $x^+ = \max(0, x)$. Generally speaking, this idea can be applied to the case of the decision functions $\hat{S}_{n,k}$ and $\tilde{S}_{n,k}$. Let us replace the log LR S_n with the GLR $\hat{S}_{n,k}$ in Eq. (28):

$$\hat{N}_r = \inf\{n \geq 1 : \hat{S}_n \geq h\};$$

$$\hat{S}_n = \begin{cases} 0 & \text{if } n_n < \tilde{m} + 1, \\ \hat{S}_{n,n-n_n+1} & \text{if } n_n \geq \tilde{m} + 1, \end{cases} \quad (29)$$

where $\hat{S}_0 = 0$, $n_0 = \tilde{m} + 1$, $n_n = \mathbf{1}_{\{(n_{n-1} < \tilde{m} + 1) \cup (S_{n-1} > 0)\}} n_{n-1} + 1$ is the counter of the observations in the current cycle of the repeated SPRT (with the GLR), $\mathbf{1}_{\{A\}}$ is the indicator of the event A : $\mathbf{1}_{\{A \text{ is false}\}} = 0$; $\mathbf{1}_{\{A \text{ is true}\}} = 1$, $\tilde{m} \geq 0$ is chosen to ensure that a MLE of θ_1 exists. When a new repeated SPRT cycle is started, we first have to check if the number of observations n_n is sufficient to ensure the existence of the MLE of θ_1 . If the number of observations n_n is less than $\tilde{m} + 1$, then we put $\hat{S}_n = 0$. This means that the current cycle of the repeated SPRT is too short and it should be continued at least until n_n becomes equal to $\tilde{m} + 1$. The indicator function $\mathbf{1}_{\{(n_{n-1} < \tilde{m} + 1) \cup (S_{n-1} > 0)\}}$ serves, first, to preserve the minimum number $\tilde{m} + 1$ of observations in the current cycle of the SPRT and, second, to finish the current cycle if $\hat{S}_{n-1} \leq 0$ and $n_{n-1} \geq \tilde{m} + 1$ (and, by this way, to restart the following one).

The behavior of the decision function \hat{S}_n and the counter n_n are shown in Fig. 5. Here, the parameter \tilde{m} is equal to 0. Before the change time $\nu = 100$ the expectation of the decision function is negative, hence, the counter n_n is stochastically small. This means that practically at every time n the detection scheme (29) restarts the SPRT between two hypothesis ($\mathcal{H}_0 : \theta = \theta_0$ vs $\mathcal{H}_1 : \theta_1 : \rho(\theta_1) = \zeta$) from scratch. After the change, the expectation of

the decision function is positive and the counter n_n grows with n (see Fig. 5). This means that the detection scheme (29) has ‘memorized’ that at the time $n - n_n + 1$ the change occurred and it carries out the observations $(Y_{n-n_n+1}, X_{n-n_n+1}), \dots, (Y_n, X_n)$. It is obvious that the change detection algorithm (29) is fully recursive, i.e. at every time n it involves only one maximization of the LR over $\theta_1 : \rho(\theta_1) = \zeta$. If we replace the GLR statistic $\hat{S}_{n,k}$ in Eq. (29) with the WLR statistic $\tilde{S}_{n,k}$, we get the other recursive stopping rule, namely, \tilde{N}_r .

3.4. ε -optimal scheme

Under certain assumptions both rules are asymptotically optimal (see [15,16,20] for details) and hold some stability when the actual value of the Kullback–Leibler information number $\zeta_l = \rho(\theta_1)$, where $\theta_1 \in \tilde{\Theta}_{1l}$ differs from the assumed one ζ_l . By choosing a desirable level ε of nonoptimality and by using the sensitivity function of the test, we get the maximum admissible difference between the actual value of the Kullback–Leibler information and the assumed one $|\zeta_l - \zeta_l|$ which leads to the definition of the zone $\tilde{\Theta}_{1l}$ (23). Typically, the maximum of nonoptimality is reached on the surfaces $\rho(\theta_1) = \delta_{l-1}$ and $\rho(\theta_1) = \delta_l$ which limit the subset $\tilde{\Theta}_{1l}$ (see Fig. 4).² Therefore, we have to choose the number L of parallel recursive tests and the optimal subdivision $\sigma = \{\zeta_1 < \zeta_2 < \dots < \zeta_L\}$ of the interval $[d_0^2/2; d_1^2/2]$ as functions of the constants d_0 , d_1 and coefficient of nonoptimality ε :

$$L = L(d_0, d_1, \varepsilon), \quad \sigma = \sigma(d_0, d_1, \varepsilon).$$

This problem will be solved in Section 4. Finally, we consider a collection of L recursive stopping rules

$$N_{\varepsilon r} = \min\{N_r(\zeta_1), \dots, N_r(\zeta_L)\}, \quad (30)$$

where $N_r(\zeta_l)$ is the stopping time (29) of the l th recursive rule (with the GLR or WLR statistics)

² Naturally, the a priori knowledge of $\rho(\theta_1)$ is important: the precisely the value of $\rho(\theta_1)$ is known, the less is the loss of optimality. Therefore, the smaller admissible value ε , the bigger the number L of parallel tests.

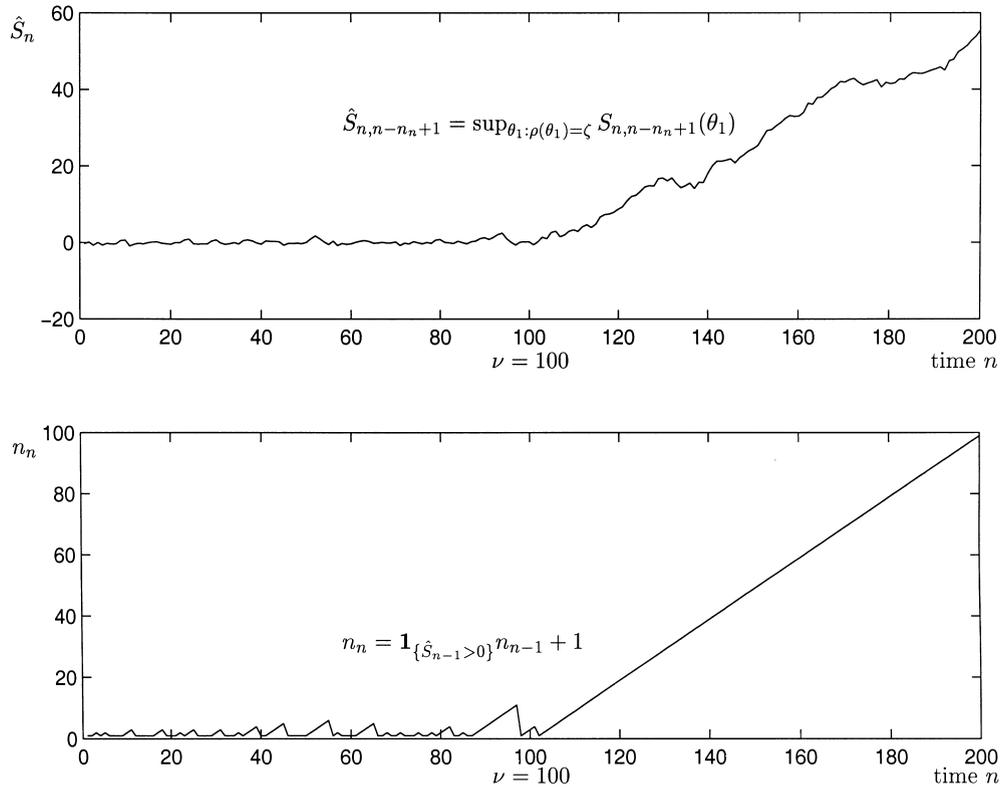


Fig. 5. The decision function \hat{S}_n and the counter n_n of the repeated SPRT (with the GLR statistic). The change time is $\nu = 100$.

designed to detect any change such that $\theta_1 \in \tilde{\Theta}_{11}$ with loss of a small part, ε , of optimality. Therefore, Eq. (30) establishes the stopping rule of the ε -optimal detection scheme.

4. Practical design of the ε -optimal algorithm

As it follows from Sections 2 and 3, the design of the detection algorithms can be done either in term of the Kullback–Leibler information number $\rho(\theta_1)$ or in term of the SNR $d = \sqrt{(\theta_1 - \theta_0)^T \Sigma^{-1} (\theta_1 - \theta_0)}$ for the first model (or $d = \sqrt{(\theta_1 - \theta_0)^T R (\theta_1 - \theta_0)}$ for the second one). In practice, it is simpler to use directly the SNR, for this reason in the rest of the paper we will consider the interval $[d_0; d_1]$ and its subdivision $\sigma = \{a_1 < a_2 < \dots < a_L\}$. We first discuss in brief two asymptotically optimal tests to detect a change

with a given SNR d (see details in [4,14–16,20]). We next design the ε -optimal scheme.

4.1. Model: $\mathcal{F}(X_n, \theta) = \theta$

4.1.1. The recursive χ^2 -GLR test

We consider the constrained GLR (24). The direct application of the method of Lagrange's multipliers (25) and the stopping rule (29) leads to the recursive χ^2 -GLR test. This test is summarized in Table 1. A detection rule which is analogous to the recursive χ^2 -GLR test was also introduced in [21].

4.1.2. The recursive χ^2 -CUSUM test

This detection rule was introduced in [14]. Its definition is analogous to the χ^2 -GLR rule. This test is summarized in Table 2, where the generalized hypergeometric function is given

Table 1
Recursive χ^2 -GLR test

Step	Action	Meaning
1.	Initialize: $n = 0, \hat{S}_0 = 0, \Sigma, d, h$	Fix initial values
2.	Iterate: $n := n + 1$ $n_n = \mathbf{1}_{\{S_{n-1} > 0\}} n_{n-1} + 1,$ $V_n = \mathbf{1}_{\{S_{n-1} > 0\}} V_{n-1} + (X_n - \theta_0),$ $\chi_n^2 = V_n^T \Sigma^{-1} V_n$ $\hat{S}_n = -n_n \frac{d^2}{2} + d \chi_n $	Take the next observation X_n Counter Cumulative sum Quadratic form (χ^2 -statistic) Decision function
3.	Check: if $\hat{S}_n \geq h$ then declare alarm: $\tilde{N}_r(d) = n$ if $\hat{S}_n < h$ then repeat step 2.	Decision rule (h is a threshold)

Table 2
Recursive χ^2 -CUSUM test

Step	Action	Meaning
1.	Initialize: $n = 0, \tilde{S}_0 = 0, \Sigma, d, h$	Fix initial values
2.	Iterate: $n := n + 1$ $n_n = \mathbf{1}_{\{S_{n-1} > 0\}} n_{n-1} + 1,$ $V_n = \mathbf{1}_{\{S_{n-1} > 0\}} V_{n-1} + (X_n - \theta_0),$ $\chi_n^2 = V_n^T \Sigma^{-1} V_n,$ $\tilde{S}_n = -n_n \frac{d^2}{2} + \log G\left(\frac{r}{2}, \frac{d^2 \chi_n^2}{4}\right)$	Take the next observation X_n Counter Cumulative sum Quadratic form (χ^2 -statistic) Decision function
3.	Check: if $\tilde{S}_n \geq h$ then declare alarm: $\tilde{N}_r(d) = n$ if $\tilde{S}_n < h$ then repeat step 2.	Decision rule (h is a threshold)

by [1]: $G(g, x) = 1 + x/g + \dots + x^n/(g(g+1)\dots(g+n-1)n!) + \dots$. In practice, the function $z \mapsto \log G(r/2, z^2/4)$ can be easily computed by applying a polynomial interpolation or, when z is large, by using the following asymptotic

approximation [1]:

$$\log G\left(\frac{r}{2}, \frac{z^2}{4}\right) = z - (r-1)\log z + \log \frac{\Gamma(r-1)}{\Gamma(r/2-1/2)} + \left(\frac{r}{2} - \frac{1}{2}\right)\log 4 + \log\left(1 + O\left(\frac{1}{z}\right)\right).$$

4.2. Model : $\mathcal{F}(X_n, \theta) = X_n^T \theta$

We consider the constrained GLR test (24). The direct application of the method of Lagrange’s multipliers (25) and the stopping rule (29) leads to the recursive constrained GLR test. The proof is given in Appendix A. This test is summarized in Table 3.

4.3. ε -optimal rule

We consider the collection (30) of L parallel recursive GLR (or CUSUM) tests given in Tables 1–3. We assume the SNR d ranges over $[d_0; d_1]$. This rule is summarized in Table 4 and the detailed explanations are given in Appendix B. The stopping time of the ε -optimal scheme is expressed as

$$N_{\varepsilon r} = \min\{N_r(a_1), N_r(a_2), \dots, N_r(a_L)\}, \quad (31)$$

where $N_r(a_l)$ is the stopping time of the recursive test designed to detect a change with the SNR a_l (see Tables 1 (or 2, 3)) and $\sigma = \{a_1 < a_2 < \dots < a_L\}$ is the optimal subdivision of the interval $[d_0; d_1]$ (B.2). As before, $\hat{N}_{\varepsilon r}$ ($\tilde{N}_{\varepsilon r}$) means that the ε -optimal rule is designed by using the GLR (CUSUM) tests. The geometric interpretations of the constrained GLR, WL GLR and ε -optimal detection scheme are given in Appendix C.

4.4. Complexity of two solutions

Let us compare the complexity of the proposed recursive ε -optimal test ($N_{\varepsilon r}$) and the WL GLR test ($\tilde{N}_{m, \tilde{m}}$). As pointed out by Lai [8], the WL GLR scheme involves $O(2 \log \gamma/d_0^2)$ LR maximizations at every time n . Hence the complexity of the WL GLR

Table 3
Recursive constrained GLR test

Step	Action	Meaning
1.	Initialize: $n = 0, r, n_0 = r + 1,$ $\hat{S}_0 = 0, R^{-1}, d, h$	Fix initial values
2.	Iterate: $n := n + 1$ $e_n = y_n - X_n^T \theta_0,$ $n_n = \mathbf{1}_{\{(n_{n-1} < r+1) \cup (S_{n-1} > 0)\}} n_{n-1} + 1,$ $V_n = \mathbf{1}_{\{(n_{n-1} < r+1) \cup (S_{n-1} > 0)\}} V_{n-1} + X_n e_n,$ if $n_n = 1$ then $P_{n-1} = R^{-1},$ $P_n = \left(I - P_{n-1} \frac{X_n X_n^T}{1 + X_n^T P_{n-1} X_n} \right) P_{n-1},$ $\hat{S}_n = \begin{cases} 0 & \text{if } n_n < r + 1 \\ -n_n \frac{d^2}{2} + d \sqrt{V_n^T n_n P_n V_n} & \text{if } n_n \geq r + 1 \end{cases}$	Take the next y_n, X_n Residual Counter Cumulative sum Re-initialization of the inverse matrix Inverse matrix
3.	Check: if $\hat{S}_n \geq h$ then declare alarm: $\hat{N}_r(d) = n$ if $\hat{S}_n < h$ then repeat step 2.	Decision rule (h is a threshold)

Table 4
Recursive ε -optimal scheme

Step	Action	Meaning
1.	Initialize: $\bar{\varepsilon}, d_0, d_1, h$ $L \text{ is the smallest integer } \geq \log \frac{d_1}{d_0} \left(\log \frac{1 + \sqrt{\bar{\varepsilon}}}{1 - \sqrt{\bar{\varepsilon}}} \right)^{-1}$ for $l = 1, \dots, L$ compute $a_l = d_0 \sqrt{\bar{\varepsilon}} \frac{(1 + \sqrt{\bar{\varepsilon}})^l}{(1 - \sqrt{\bar{\varepsilon}})^{l-1}},$ initialize L parallel tests	Tuning parameters Number of parallel tests Optimal subdivision of $[d_0; d_1]$ See Table 1 (or 2,3)
2.	Iterate: $n := n + 1$ for $l = 1, \dots, L$ compute $\hat{S}_n(a_l)$ (or $\tilde{S}_n(a_l)$)	Take the next observation L parallel tests, See Table 1 (or 2,3)
3.	Check: if $\max\{\hat{S}_n(a_1), \dots, \hat{S}_n(a_L)\} \geq h$ then declare alarm: $\hat{N}_{er} = n$ (or $\tilde{N}_{er} = n$) if $\max\{\hat{S}_n(a_1), \dots, \hat{S}_n(a_L)\} < h$ then repeat step 2.	Decision rule (h is a threshold)

scheme is $O(2 \log \gamma / d_0^2)$. The proposed ε -optimal test involves L LR maximizations at every time n , therefore, its complexity is L . The comparison of the efficiency versus complexity for the ε -optimal scheme and the WL GLR test is

summarized in Table 5. Let us compute the following ratio:

$$\frac{\text{Complexity of } \hat{N}_{m, \bar{m}}}{\text{Complexity of } N_{er}} = \frac{2 \log \gamma}{L d_0^2},$$

Table 5
Comparison of the efficiency versus complexity for the ε -optimal and WL GLR tests as $\gamma \rightarrow \infty$

Test	Loss of optimality	Complexity (number of LR maximizations)
ε -optimal	ε	The smallest integer $\geq \log \frac{d_1}{d_0} \left(\log \frac{1 + \sqrt{\varepsilon}}{1 - \sqrt{\varepsilon}} \right)^{-1}$
WL GLR (optimal)	0	$O\left(\frac{2 \log \gamma}{d_0^2}\right)$

as a function of the mean time before a false alarm γ for $d_0 = 0.3$, $d_1 = 10$, and $\bar{\varepsilon} = 0.05; 0.1; 0.2; 0.4$. These results are shown in Fig. 6.

5. Comparison against the WL GLR scheme: simulation results

5.1. Simulation model

The goal of this section is twofold. First, we compare the statistical performances and the

computational complexity of the recursive ε -optimal and WL GLR tests for the problem of detecting changes in the regression model $y_n = X_n^T \theta + \xi_n$ by using the Monte-Carlo simulation. Second, we use the simulation results to confirm the theoretical performances of the ε -optimal test. A simulation comparison of the ε -optimal scheme with the WL GLR scheme for the additive model (5) has been discussed earlier [20]. We assume now that the inputs X_n are zero mean gaussian random vectors with covariance matrix R . The parameters have been chosen in the following manner: $r = 5$, $\theta_0^T = (0, \dots, 0)$, $\xi_n \sim \mathcal{N}(0,1)$,

$$R_5 = \begin{pmatrix} 1 & 0.8 & 0.2 & 0.1 & -0.2 \\ 0.8 & 1 & -0.2 & 0.1 & 0.1 \\ 0.2 & -0.2 & 1 & 0.5 & 0.1 \\ 0.1 & 0.1 & 0.5 & 1 & 0.3 \\ -0.2 & 0.1 & 0.1 & 0.3 & 1 \end{pmatrix},$$

$$\det R_5 = 0.062.$$

We assume that the parameter θ_1 is given by the equation $\theta_1^T R_5 \theta_1 = d^2$, where the SNR d varies

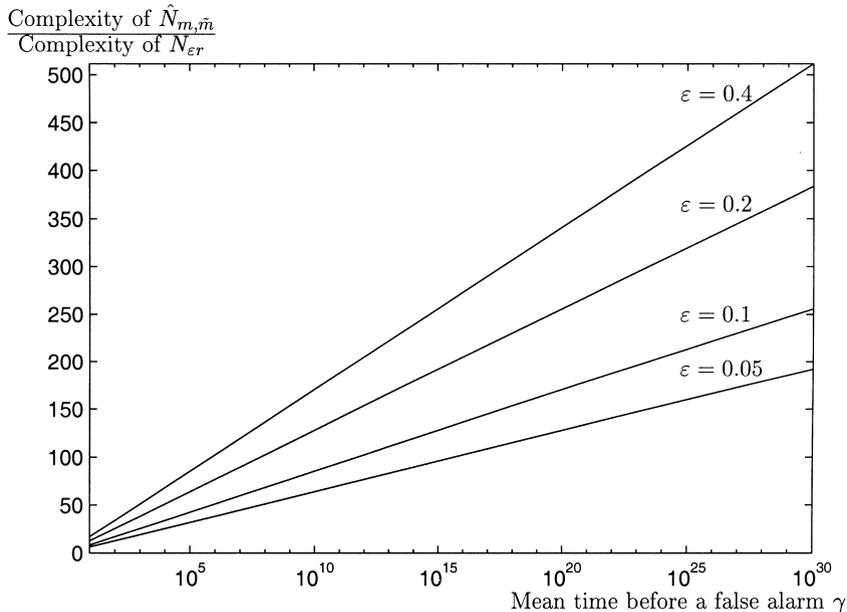


Fig. 6. Comparison of the efficiency versus complexity for the ε -optimal and WL GLR tests.

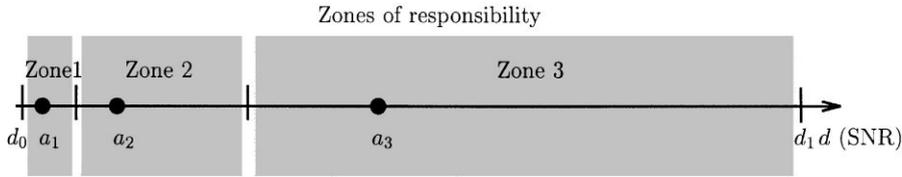


Fig. 7. The subdivision $\sigma = \{a_1 < a_2 < a_3\}$ of the interval $[d_0; d_1]$ and three zones of responsibility.

between $d_0 = 0.3$ and $d_1 = 10$. We fix the coefficient of nonoptimality as $\bar{\varepsilon} = 0.3$ and the mean time before a false alarm as $\gamma = 10^4$. It follows from Eq. (B.2) that it is enough to run three parallel χ^2 -GLR tests to get the level $\bar{\varepsilon} = 0.3$ of nonoptimality and that the assumed values of the SNR are $a_1 = 0.464$, $a_2 = 1.589$ and $a_3 = 5.437$. The zones of responsibility $[0.3, 1.027]$, $[1.027, 3.513]$ and $[3.513, 12.022]$ are shown by shaded regions in Fig. 7.

5.2. Statistical performances

The Monte-Carlo simulation was organized in the following manner. The ε -optimal recursive algorithms $\hat{N}_{\varepsilon r}$ summarized in Tables 3 and 4 have been compared against the WL GLR test given in

Tables 6. The threshold value h in the ε -optimal algorithm (31), (A.1), (A.2) or in the WL GLR algorithm (A.3) was so chosen that $\mathbb{E}^\infty(\hat{N}_{\varepsilon r}) = \mathbb{E}^\infty(\hat{N}_{m,r}) = 10^4$. To evaluate h we have used an iterative scheme. Let us define the following function $f: \mathbb{R} \rightarrow \mathbb{R}$, $f(h) = \mathbb{E}^\infty(N; h) - \gamma$, where $\gamma = 10^4$. The secant method for the solution of equation $f(h) = 0$ was used:

$$h_{n+1} = h_n - f(h_n) \frac{h_n - h_{n-1}}{f(h_n) - f(h_{n-1})},$$

$$h_0 = \log \gamma + \mu, h_{-1} = \log \gamma, \text{ with } \mu = 1.$$

At each iteration n , 10^3 independent repetitions of the Monte-Carlo simulation were performed to estimate $f(h_n)$. The iterative process was stopped

Table 6
WL GLR test

Step	Action	Meaning
1.	Initialize: $n = 0, r, m: m > r, R^{-1}, h$	Fix initial values
2.	Iterate: $n = n + 1$ $e_n = y_n - X_n^T \theta_0$, if $n \leq r$ then repeat step 2, if $n > r$ that execute step 2.1	Take the next y_n, X_n Residual Wait for the next observation First start of the inside cycle
2.1	Initialize $V_{n+1} = 0, P_{n,n+1} = R^{-1}$, $l = \max\{0, n - m\} + 1$,	New time window $[l; n]$
2.2	Inside cycle for iterations $k = n, n - 1, \dots, l$	Iterations in the backward direction
	$P_{n,k} = \left(I - P_{n,k+1} \frac{X_n X_n^T}{1 + X_n^T P_{n,k+1} X_n} \right) P_{n,k+1},$	Inverse matrix
	$V_k = V_{k+1} + X_k e_k,$ if $k \leq n - r$ then $\hat{S}_{n,k} = \frac{1}{2} V_k^T P_{n,k} V_k$	Cumulative sum Decision function
3.	Check: if $\max\{\hat{S}_{n,n-r}, \dots, \hat{S}_{n,l}\} \geq h$ then declare alarm : $\hat{N}_{m,r} = n$ if $\max\{\hat{S}_{n,n-r}, \dots, \hat{S}_{n,l}\} < h$ then repeat step 2.	Decision rule (h is a threshold)

when the estimated value of $\mathbb{E}^\infty(\hat{N}_{sr})$ (or $\mathbb{E}^\infty(\hat{N}_{m,r})$) differs from 10^4 by no more than 50 (0.5% of 10^4). In this manner the threshold h was set to be $h = 10.98$ for the ε -optimal algorithm and was set to be $h = 13.3$ for the WL GLR algorithm. The parameter m in the WL GLR algorithm (see Table 6) was set to be $m = 400$ (see explanations below). Until the change time v the parameter θ is equal to θ_0 , hence, the observations y_1, y_2, \dots, y_{v-1} were generated by the gaussian distribution $\mathcal{N}(0,1)$. Beginning from the change time v the observations y_v, y_{v+1}, \dots were generated by the equation $y_n = X_n^T \theta_1 + \xi_n$. To evaluate $\mathbb{E}(\hat{N}_{0.3r}; d, 10^4)$ and $\mathbb{E}(\hat{N}_{400,5}; d, 10^4)$, 10^3 repetitions of the detection procedures were performed for each value of $v = 1, 2, 3, \dots, 100$ and $d \in [0.3; 10]$. The simulation shows that the variation of the estimate mean time delay for detection is negligible (with respect its standard deviation) when $v > 100$. Hence, we estimate the mean time delay for detection by replacing the value $\mathbb{E}(N) = \sup_{v \geq 1} \mathbb{E}_d^v(N - v + 1 | N \geq v)$ by its Monte-Carlo estimate

$$\sup_{v \geq 1} \mathbb{E}_d^v(N - v + 1 | N \geq v) \simeq \max_{1 \leq v \leq 100} \left\{ \frac{1}{n_d} \sum_{i=1}^{10^3} (\hat{N}_i - v + 1) \mathbf{1}_{\{\hat{N}_i \geq v\}} \right\},$$

where \hat{N}_i is the time of detection in the i th statistical experiment and $n_d = \sum_{i=1}^{10^3} \mathbf{1}_{\{\hat{N}_i \geq v\}}$. Fig. 8 reports the results of a simulation study of the statistical performances of the above ε -optimal and WL GLR tests. Both, the simulated and theoretic $\mathbb{E}(N; d, 10^4)$ as functions of the SNR d for these tests are presented here. Because the WL GLR test is asymptotically optimal, we use (21) as a theoretic expression for the mean time delay for detection, on the other hand we use (B.4) as a theoretic mean time delay for the ε -optimal test. The results of a simulation study (see Fig. 8) completely confirm the theoretic performances of the ε -optimal test, it even performs better than it can be expected from (B.4).

5.3. Computational complexity

The complexity of the 0.3-optimal test is three LR maximizations at every time. It is difficult to use directly the asymptotic equation $m = O((2 \log \gamma / d_0^2))$ for the WL GLR test because the adequate choice of the parameter m implicitly involves the higher moments of the stopping time $\hat{N}_{m,r}$. We have first chosen $m = (2 \log 10^4 / 0.3^2) \simeq 206$. The simulation showed that this choice leads to an underestimate of the statistical properties of the WL GLR test. To be sure that the comparison is correct, we have

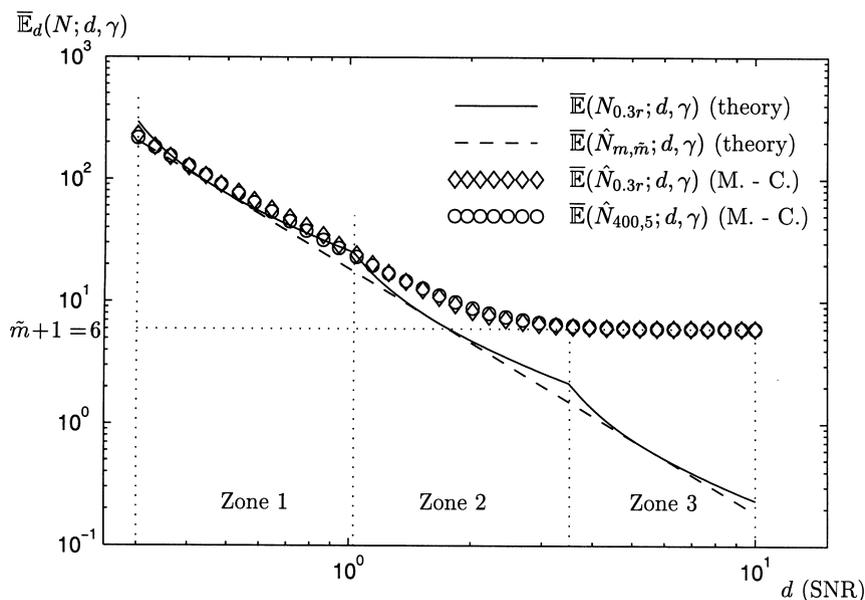


Fig. 8. The mean time detection delays of the WLGLR and ε -optimal tests.

next repeated the simulation with the $m = 400 > (2 \log 10^4 / 0.3^2)$. Therefore, the comparative computational complexity of the ε -optimal test $\hat{N}_{0.3r}$ and the WL GLR test $\hat{N}_{400,5}$ is presented by the following ratio:

$$\frac{\text{Complexity of } \hat{N}_{400,5}}{\text{Complexity of } \hat{N}_{0.3r}} = \frac{395}{3} \simeq 132.$$

6. Conclusions

The detection of abrupt changes in the model $Y_n = \mathcal{F}(X_n, \theta) + \zeta_n$ is discussed. Two particular cases have been considered: (1) $\mathcal{F}(X_n, \theta) = \theta$; (2) $\mathcal{F}(X_n, \theta) = X_n^T \theta$. The parameter θ is assumed to be unknown after the change. A simple ε -optimal recursive scheme is proposed to solve the problem. A method for tuning the parameters of this ε -optimal detection scheme is given (see Eqs. (B.1)–(B.4)). By using this method, the designer can easily find a trade-off between the complexity of the proposed change detection algorithm and its efficiency. The proposed scheme has been compared with the asymptotically optimal WL GLR scheme which is usually used to solve such a problem.

Appendix A. Proof of the GLR tests

A.1. Recursive constrained GLR test

We consider the constrained GLR test (24) and the regression model (8). To simplify the notations we apply the transformation $e_n = y_n - X_n^T \theta_0$ to the output y_n . This leads the following representation of model (8):

$$e_n = \begin{cases} \zeta_n & \text{if } n < v, \\ X_n^T \beta + \zeta_n & \text{if } n \geq v, \end{cases}$$

where $\beta = \theta_1 - \theta_0$. It follows from Eq. (24) that

$$\hat{S}_{n,k} = \max_{\beta: 2\rho(\beta)=d^2} \left\{ \sum_{i=k}^n \beta^T X_i e_i - \frac{1}{2} \beta^T \sum_{i=k}^n X_i X_i^T \beta \right\},$$

where $2\rho(\beta) = \beta^T 1 / (n - k + 1) \sum_{i=k}^n X_i X_i^T \beta$. The application of the method of Lagrange's multipliers

(25) leads to the following results:

$$\begin{aligned} \frac{\partial S_{n,k}(\beta)}{\partial \beta} + 2\lambda \frac{\partial \rho(\beta)}{\partial \beta} \\ = \sum_{i=k}^n X_i e_i - \left(1 - \frac{2\lambda}{n-k+1} \right) \sum_{i=k}^n X_i X_i^T \beta = 0, \\ 2\rho(\beta) = \beta^T \frac{1}{n-k+1} \sum_{i=k}^n X_i X_i^T \beta = d^2. \end{aligned}$$

If the matrix $\sum_{i=k}^n X_i X_i^T$ is nonsingular, then there is a unique solution for the above system. It leads to the following expression of the GLR:

$$\begin{aligned} \hat{S}_{n,k} &= -(n-k+1) \frac{d^2}{2} + d |\chi_{n,k}|, \\ \chi_{n,k}^2 &= \sum_{i=k}^n X_i^T e_i \left(\frac{1}{n-k+1} \sum_{i=k}^n X_i X_i^T \right)^{-1} \sum_{i=k}^n X_i e_i. \end{aligned}$$

Taking into account the recursive method of computing the inverse of $\sum_{i=k}^n X_i X_i^T$ knowing $\sum_{i=k}^{n-1} X_i X_i^T$ and the fact that $\tilde{m} = r$, we get the recursive GLR test:

$$\hat{N}_r(d) = \inf\{n \geq 1: \hat{S}_n \geq h\}, \quad (\text{A.1})$$

$$\hat{S}_n = \begin{cases} 0 & \text{if } n_n < r + 1, \\ -n_n \frac{d^2}{2} + d \sqrt{V_n^T n_n P_n V_n} & \text{if } n_n \geq r + 1, \end{cases}$$

$$P_n = \left(I - P_{n-1} \frac{X_n X_n^T}{1 + X_n^T P_{n-1} X_n} \right) P_{n-1}$$

$$\text{with } P_{n-1} = P_0 \text{ if } n_n = 1, \quad (\text{A.2})$$

where $P_n \triangleq (\sum_{i=n-n_n+1}^n X_i X_i^T)^{-1}$, $V_n = \mathbf{1}_{\{(n_{n-1} < r+1) \cup (S_{n-1} > 0)\}} V_{n-1} + X_n e_n$ and $n_n = \mathbf{1}_{\{(n_{n-1} < r+1) \cup (S_{n-1} > 0)\}} n_{n-1} + 1$. The initial conditions are $\hat{S}_0 = 0$, $n_0 = r + 1$ and $P_0 = R^{-1}$. In practice, the matrix P_0 can be also chosen as ωI , where $\omega = 10^2 - 10^5$.

A.2. WL GLR test

Let us briefly derive the WL GLR test for the model $\mathcal{F}(X_n, \theta) = X_n^T \theta$. The details can be found in [10]. It follows from Eq. (3) that the stopping time of the WL GLR test is given by

$$\begin{aligned} \hat{N}_{m,r} = \inf \left\{ n > r: \max_{\max\{0, n-m\} + 1 \leq k \leq n-r} \frac{1}{2} \sum_{i=k}^n X_i^T e_i \right. \\ \left. \cdot \left(\sum_{i=k}^n X_i X_i^T \right)^{-1} \sum_{i=k}^n X_i e_i \geq h \right\}, \quad (\text{A.3}) \end{aligned}$$

where $1 \leq r < m$, $e_n = y_n - X_n^T \theta_0$ and the inverse matrix $P_{n,k} \triangleq (\sum_{i=k}^n X_i X_i^T)^{-1}$ is calculated by using the recursive method (A.2) applied in the backward direction ($k = n, n-1, n-2, \dots$):

$$P_{n,k} = \left(I - P_{n,k+1} \frac{X_n X_n^T}{1 + X_n^T P_{n,k+1} X_n} \right) P_{n,k+1}$$

with $P_{n,n+1} = R^{-1}$ if $k = n$.

Appendix B. Properties of the proposed tests

B.1. Model: $\mathcal{F}(X_n, \theta) = \theta$

It follows from [15,20] that the asymptotic relation between the mean time delay for detection and the mean time before a false alarm for the ε -optimal rule is given by

$$\bar{\mathbb{E}}(N_{er}; d, \gamma) \lesssim \frac{2 \log \gamma}{d^2 - (d - a_{l_0})^2} \quad \text{as } \gamma \rightarrow \infty, \quad (\text{B.1})$$

where $d \in [d_0; d_1]$, $l_0 = \operatorname{argmin}\{|d - a_l| \mid l = 1, \dots, L\}$ and $\sigma = \{a_1 < a_2 < \dots < a_L\}$ is a subdivision of the interval $[d_0; d_1]$ such that $d_0 \leq a_1 < 2d_0$ and $a_L \leq d_1$. Let us discuss Eq. (B.1). The ε -optimal rule (31) is a collection of L parallel recursive tests. The stopping (alarm) time is set at the first instant N_{er} for which at least one decision function reaches the threshold h . The bound between two neighboring zones of responsibility (say, the zones l and $l+1$) is the center of the interval $[a_l, a_{l+1}]$. The zones of responsibility are shown by shaded regions and the bounds between neighboring zones are shown by vertical lines in Fig. 7. Let us assume that the actual value d of the SNR belongs to the l_0 th zone. It is easy to show that the smallest mean time delay for detection corresponds to the l_0 th recursive test. This smallest mean time delay for detection is an upper bound for the mean time delay $\bar{\mathbb{E}}(N_{er}; d, \gamma)$ of the ε -optimal rule. This fact explains Eq. (B.1) and the definition of the number l_0 . Let us fix some acceptable value $\bar{\varepsilon}$ of the coefficient of *nonoptimality* (see Eq. (22)), the number L and the subdivision σ should be chosen in the

following manner:

$$L \text{ is the smallest integer } \geq \log \frac{d_1}{d_0} \left(\log \frac{1 + \sqrt{\bar{\varepsilon}}}{1 - \sqrt{\bar{\varepsilon}}} \right)^{-1} \quad (\text{B.2})$$

and

$$a_l = d_0 \sqrt{\bar{\varepsilon}} \frac{(1 + \sqrt{\bar{\varepsilon}})^l}{(1 - \sqrt{\bar{\varepsilon}})^{l-1}}, \quad l = 1, 2, \dots, L.$$

Because Eq. (B.1) has an asymptotic character, it is not very accurate for large values of SNR. Obviously the detection delay is greater than or equal to $\tilde{m} + 1$. In the case of model $\mathcal{F}(X_n, \theta) = \theta$ the parameter \tilde{m} is equal to 0. This leads to the following heuristic modification of Eq. (B.1):

$$\bar{\mathbb{E}}(N_{er}; d, \gamma) \lesssim \max \left\{ 1, \frac{2 \log \gamma}{d^2 - (d - a_{l_0})^2} \right\}. \quad (\text{B.3})$$

This equation can be used to estimate the ARL function. It follows from Eq. (18) that $\text{ARL}(0) = \mathbb{E}_{\theta_0}^1(N) = \mathbb{E}^\infty(N)$, hence $\text{ARL}(0) = \gamma$. On the other hand $\text{ARL}(d) = \bar{\mathbb{E}}(N_{er}; d, \gamma)$.

B.2. Model: $\mathcal{F}(X_n, \theta) = X_n^T \theta$

The optimal detection of abrupt changes in this model by using Lorden's criterion [11] and/or Eqs. (19)–(20) has been discussed by Bansal and Papanтони-Kazakos [2], Yao [30] and recently by Lai [8,9]. It has been shown that the lower bound $n(d, \gamma) \triangleq \inf\{\bar{\mathbb{E}}(N; d, \gamma) : N \in \mathcal{H}_\gamma\} \sim (2 \log \gamma / d^2)$ is also valid for the regression model as $\gamma \rightarrow \infty$. The goal of this paper is to discuss the methodological and practical aspects of the ε -optimal scheme, for this reason we omit mathematical details. We assume that the mean time delay for detection given by Eq. (B.3) is valid also for the regression model $\mathcal{F}(X_n, \theta) = X_n^T \theta$:

$$\bar{\mathbb{E}}(N_{er}; d, \gamma) \lesssim \max \left\{ r + 1, \frac{2 \log \gamma}{d^2 - (d - a_{l_0})^2} \right\}. \quad (\text{B.4})$$

To find the number L and the subdivision σ we apply again Eq. (B.2). The results of Monte-Carlo simulation show that this approach can be used at least as an asymptotic approximation.

B.3. Dynamic profile after change

Let us assume now that the parameter θ changes not abruptly (see Eqs. (5) and (8)) but smoothly, namely $\theta_1 = \theta_1(n - v + 1)$, $n \geq v$, is a function of time. We assume that the dynamic profile $\theta_1(n - v + 1)$ is unknown a priori and, hence, it cannot be directly included in the log LR $S_{n,k}(\theta_1) = \sum_{j=k}^n \log(\varphi_{\theta_1(n-k+1)}(Y_j|X_j)/(\varphi_{\theta_0}(Y_j|X_j)))$ as it is proposed in [5], where the problem of time-varying change detection is considered for the scalar case. We consider here that the unknown dynamic profile $\theta_1(n - v + 1)$ is a nuisance parameter. We briefly discuss now its impact on the performance index (mean time delay for detection $\bar{\mathbb{E}}(N, \gamma)$) of the considered detection algorithms. It is worth noting that to our knowledge no result in a mathematically precise sense exists in the literature, for this reason our discussion is heuristic. There are two different aspects of this problem: (i) the impact of the dynamic profile $\theta_1 = \theta_1(n - v + 1)$ on $\bar{\mathbb{E}}(N, \gamma)$ via the MLE estimate of θ_1 ; (ii) the impact of the dynamic profile $\theta_1 = \theta_1(n - v + 1)$ via the actual value of the Kullback–Leibler information number $\zeta = \rho(\theta_1)$. The impact via the MLE estimate of θ_1 is important only for the model $\mathcal{F}(X_n, \theta) = X_n^T \theta$, because for the model $\mathcal{F}(X_n, \theta) = \theta$ the MLE is instantaneous (based on one observation) and, hence, only the second aspect is important. For the model $\mathcal{F}(X_n, \theta) = X_n^T \theta$, the dynamic profile $\theta_1 = \theta_1(n - v + 1)$ produces a negative effect on the MLE estimate of θ_1 . This effect is very difficult to estimate mathematically and the Monte-Carlo simulation method can be recommended in this case. For the model $\mathcal{F}(X_n, \theta) = \theta$, the situation is a bit simpler. It will affect the mean time delay for detection $\bar{\mathbb{E}}(N, \gamma)$, but the following heuristic role can be proposed in such a case: (i) to compute the dynamic profile of the Kullback–Leibler information number $\zeta = \zeta(n - v + 1)$; (ii) to solve (numerically) the following asymptotic equation [17] $\sum_{n=1}^{\tau} \zeta(n) = \log \gamma$, where $\tau \simeq \bar{\mathbb{E}}(N, \gamma)$.

B.4. Some alternative solutions

Let us briefly discuss some alternative solutions to the problem of detecting changes in the models

$\mathcal{F}(X_n, \theta) = \theta$ and $\mathcal{F}(X_n, \theta) = X_n^T \theta$. For the model $\mathcal{F}(X_n, \theta) = \theta$ with a known value of the SNR $d = \sqrt{(\theta_1 - \theta_0)^T \Sigma^{-1} (\theta_1 - \theta_0)}$, a fixed sample size (FSS) test with nonoverlapping blocks of m observations is compared against the optimum sequential tests $\hat{N}(d)$ and $\tilde{N}(d)$ (χ^2 -GLR or CUSUM, see Tables 1 and 2) in [19]. The stopping time of this FSS test is given by $\bar{N}(d) = \inf_{j \geq 1} \{mj : |\chi_{(j-1)m+1}^m| \geq mh\}$, where $(\chi_k^n)^2 = (\bar{V}_k^n)^T \Sigma^{-1} \bar{V}_k^n$ and $\bar{V}_k^n = \sum_{i=k}^n (X_i - \theta_0)$. A more conservative performance criterion of Lorden [11] $\bar{\mathbb{E}}^*(N) = \sup_{v \geq 1} \text{esssup} \mathbb{E}_v(N - v + 1 | N \geq v, X_1, \dots, X_{v-1})$ has been used for this comparison. It is shown that

$$\begin{aligned} \bar{\mathbb{E}}^*(\hat{N}; d, \gamma) &\sim \bar{\mathbb{E}}^*(\tilde{N}; d, \gamma) \sim \frac{2 \log \gamma}{d^2} \\ &\lesssim \bar{\mathbb{E}}^*(\bar{N}; d, \gamma) \lesssim \frac{4 \log \gamma}{d^2} \quad \text{as } \gamma \rightarrow \infty \end{aligned}$$

and (by numerical method) that the true value of $\bar{\mathbb{E}}^*(\bar{N}; \gamma)$ is very close to its upper bound, hence, the χ^2 -CUSUM (or GLR) test is asymptotically twice as good as the FSS test. Moreover, it is shown that the optimal sample size is $m \sim 2 \log \gamma / d^2$ and, hence, if the SNR d is unknown then an additional loss of performance can be expected in the case of the FSS test. For the model $\mathcal{F}(X_n, \theta) = X_n^T \theta$, the above results are applicable as approximation. The “moving window” FSS, nonoverlapping FSS tests and the window-limited CUSUM test are compared against the WL GLR test $\hat{N}_{m,m}$ (3) by Monte Carlo simulations in [9,10]. The state space model with an additive change is used for this comparison. The conclusion is the following: if the parameter m is so chosen that $m > \log \gamma / \rho_{\min}$ with $\rho_{\min} = \inf_{\theta_1 \in \Theta_1} \rho(\theta_1)$, then the WL GLR test realizes the best results (even for a non-asymptotic value $\gamma \simeq 10^3$). If m is misspecified ($m < \log \gamma / \rho_{\min}$), then there are the values of θ_1 when the alternative tests perform better. This fact emphasizes the importance of a correct choice of the parameter m . Therefore, the WL GLR test with a conveniently chosen parameter m is a good benchmark, for this reason the proposed ε -optimal test is compared with the WL GLR test in Section 5.

Appendix C. Geometric interpretations

Let us discuss now the geometric interpretations of the GLR, WL GLR and ε -optimal detection schemes. These interpretations are especially simple in the case of the model $\mathcal{F}(X_n, \theta) = \theta$ (see Eq. (5)), where $\theta \in \mathbb{R}^2$ and covariance matrix is scalar $\Sigma = \sigma^2 I$, i.e. for the bivariate detection scheme. By substitution of the log LR

$$S_{n,k}(\theta_1) = \sum_{j=k}^n \log \frac{\varphi_{\theta_1}(Y_j)}{\varphi_{\theta_0}(Y_j)}$$

$$= \frac{n-k+1}{2\sigma^2} (-\|\bar{Y}_{n,k} - \theta_1\|^2 + \|\bar{Y}_{n,k} - \theta_0\|^2)$$

with $\bar{Y}_{n,k} = 1/(n-k+1) \sum_{j=k}^n Y_j$ in Eq. (2) we get the stopping time of the GLR scheme:

$$\hat{N} = \inf \left\{ n \geq 1 : \max_{1 \leq k \leq n} \sup_{\theta_1 \in \mathbb{R}^2} S_{n,k}(\theta_1) \geq h \right\}$$

with

$$\sup_{\theta_1 \in \mathbb{R}^2} S_{n,k}(\theta_1) = \frac{1}{2\sigma^2(n-k+1)} \left\| \sum_{i=k}^n (Y_i - \theta_0) \right\|^2.$$

The graphical solution of the GLR test consists in plotting the bivariate cumulative sum $Z_n = (z_n^1, z_n^2)^T = \sum_{i=1}^n (Y_i - \theta_0)$ versus n and placing the vertex of the stopping surface of revolution (this is a paraboloid of revolution) on the latest point M with coordinates (z_n^1, n, z_n^2) (see Fig. 9, upper row). Fig. 9, upper row, shows the GLR stopping surface when the variance σ^2 is equal to 1 and

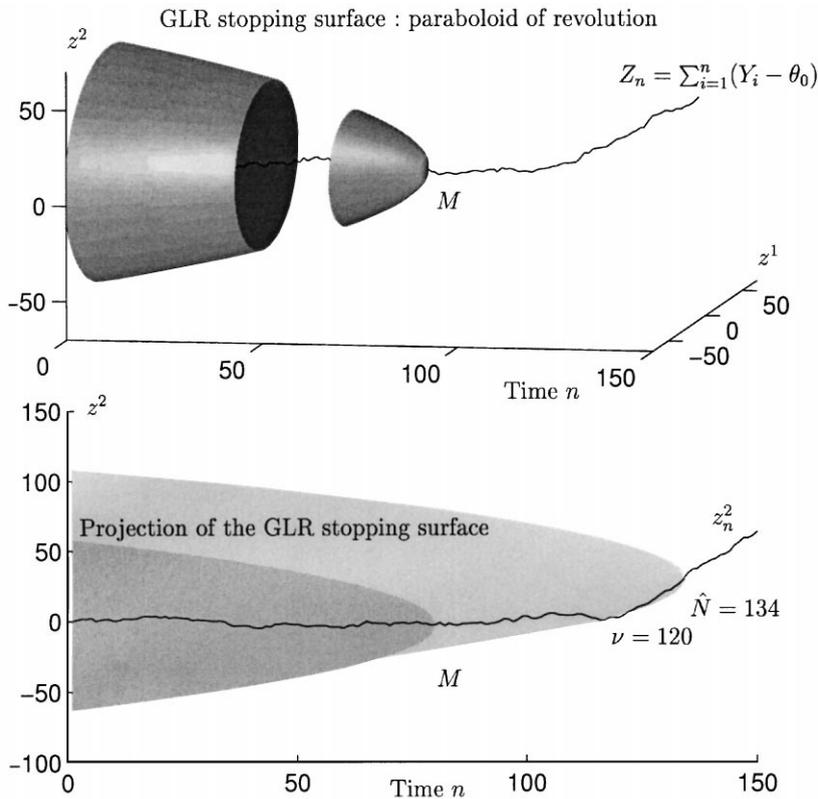


Fig. 9. Geometric interpretation of the GLR stopping bounds.

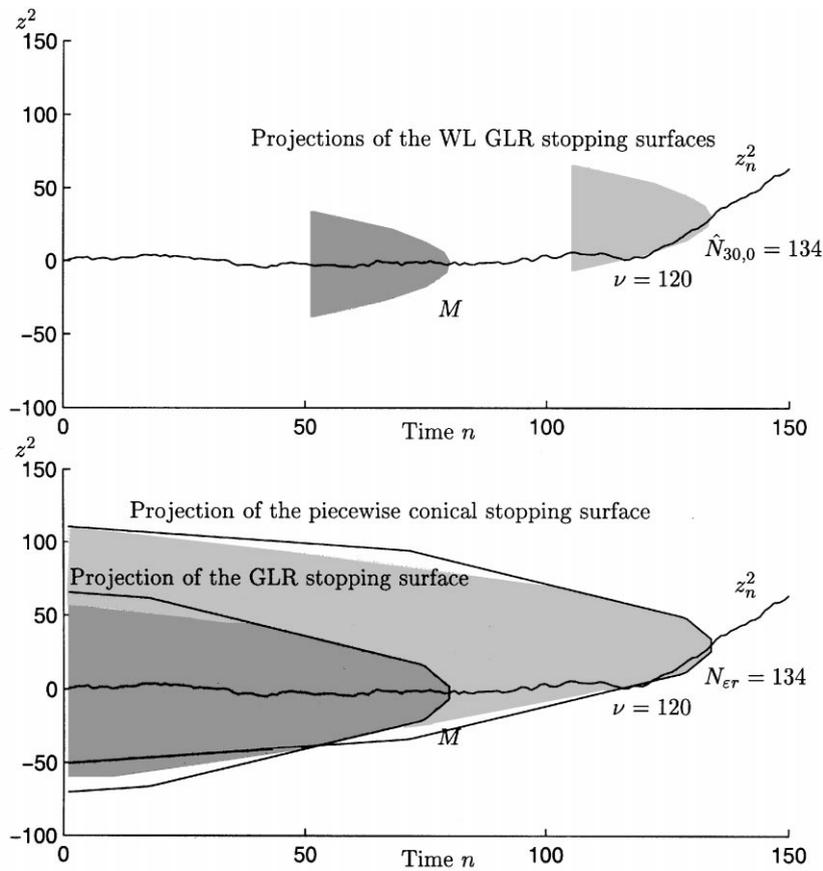


Fig. 10. Geometric interpretation of the WL GLR and ϵ -optimal stopping bounds.

the threshold h is equal to 23. The decision rule \hat{N} consists in stopping at the first time n such that a previously plotted point of the bivariate cumulative sum $\sum_{i=k}^n (Y_i - \theta_0)$ crosses the stopping surface of revolution with the generatrix $r_l = \sqrt{2\sigma^2 l h}$, where $l = 0, 1, 2, \dots, n$. This situation is shown in Fig. 9, low row, by projecting the paraboloid on the z^2-n plane, see shaded parabolic regions. The graphical solution of the WL GLR test consists in the same manipulations but the stopping surface now is a truncated paraboloid of revolution with the generatrix $r_l = \sqrt{2\sigma^2 l h}$, where $l = 0, 1, 2, \dots, m$. This solution is shown in Fig. 10, upper row, when the parameter m is equal to 30, by projecting the truncated paraboloid on the z^2-n plane. The above graphical rule “stop if a previously plotted point lies on the opposite side of the stopping surface” is

also valid in the case of the ϵ -optimal test. It follows from [15,20] that the recursive χ^2 -GLR (or χ^2 -CUSUM) tests designed to detect a change with the SNR d and given by Tables 1 and 2 are asymptotically equivalent to the following nonrecursive constrained GLR test:

$$N(d) = \inf \left\{ n \geq 1 : \max_{1 \leq k \leq n} \sup_{\theta_1 : \|\theta_1 - \theta_0\| = d\sigma} S_{n,k}(\theta_1) \geq h \right\}. \tag{C.1}$$

As it follows from Eqs. (24) and (25), the constrained GLR is given by

$$\begin{aligned} & \sup_{\theta_1 : \|\theta_1 - \theta_0\| = d\sigma} S_{n,k}(\theta_1) \\ &= -\frac{n-k+1}{2} d^2 + \frac{d}{\sigma} \left\| \sum_{i=k}^n (Y_i - \theta_0) \right\|. \end{aligned}$$

Therefore, now the stopping bound is a conical surface with the generatrix $\hat{r}_l = h\sigma/d + d\sigma l/2$, where $l = 1, 2, \dots, n$. Let us establish the relation between the generatrix r_l of the GLR test and the generatrix \hat{r}_l of the constrained GLR test. To do this, we replace the integer index l by the continuous variable x , hence, $r(x) = \sqrt{2\sigma^2 hx}$ and $\hat{r}(x) = h\sigma/d + d\sigma x/2$. It follows from [15,20] that the asymptotic mean time delay for the constrained GLR is given by $2h/d^2$. Let us consider $x_0 = 2h/d^2$. It is easy to show that the graph of the linear function $x \mapsto \hat{r}(x)$ is a tangent line to the curve given by the equation $r(x) = \sqrt{2\sigma^2 hx}$ at the point $(x_0, r(x_0))$. The ε -optimal test with the stopping time N_{ε} (31) is a collection of L parallel recursive χ^2 -GLR (or χ^2 -CUSUM) tests with specially chosen values a_1, \dots, a_L . This test is asymptotically equivalent to the collection $N_{\varepsilon} = \{N(a_1), \dots, N(a_L)\}$ of the nonrecursive stopping rules (C.1). It follows from the definition of the stopping time N_{ε} that the generatrix of the corresponding stopping surface is a piecewise linear function (see Fig. 10, lower row). These pieces given by the equations $\hat{r}_i(x) = h\sigma/a_i + a_i\sigma x/2$, $i = 1, \dots, L$ are tangent lines to the generatrix of the GLR stopping surface given by the equation $r(x) = \sqrt{2\sigma^2 hx}$ at the points $(x_i, r(x_i))$, where $x_i = 2h/a_i^2$. Therefore, the graphical solution of the ε -optimal test looks like a piecewise conical approximation from outside for the stopping surface (paraboloid) of the GLR test (see the projection of these stopping surfaces on the z^2 - n plane in Fig. 10, lower row). The portion $\hat{r}_i(x) = h\sigma/a_i + a_i\sigma x/2$ of this piecewise conical approximation corresponds to the zone of responsibility of the test $\hat{N}(a_i)$ (see Fig. 4). Fig. 10, low row, shows the approximation of the GLR scheme by three parallel tests when the assumed values of the SNR are $a_1 = 0.464$, $a_2 = 1.589$ and $a_3 = 5.437$. Here, the shaded parabolic regions represent the projections of the GLR test stopping surface and the piecewise linear bounds represent the projections of the ε -optimal test stopping surface.

For further reading

The following reference is also of interest to the reader: [24].

References

- [1] M. Abramowitz, I.A. Stegun (Eds.), Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, Applied Mathematics Series vol. 55, United States Department of Commerce, National Bureau of Standards, 1964.
- [2] R.K. Bansal, P. Papantoni-Kazakos, An algorithm for detecting a change in a stochastic process, IEEE Trans. Inform. Theory 32 (March 1986) 227–235.
- [3] M. Basseville, Information criteria for residual generation and fault detection and isolation, Automatica 33 (5) (1997) 783–803.
- [4] M. Basseville, I.V. Nikiforov, Detection of Abrupt Changes. Theory and applications, Information and System Sciences Series, Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [5] S.D. Blostein, Quickest detection of a time-varying change in distribution, IEEE Trans. Inform. Theory 37 (4) (1991) 1116–1122.
- [6] A.A. Borovkov, Theory of Mathematical Statistics – Estimation and Hypotheses Testing, Naouka, Moscow, 1984 (in Russian) (translated into English under the title: Mathematical Statistics, Gordon and Breach Science Publishers, London, 1998).
- [7] D.H. Brandwood, A complex gradient operator and its application in adaptive array theory, IEE Proc. Parts F and H 130 (1) (February 1983) 11–16.
- [8] T.L. Lai, Sequential changepoint detection in quality control and dynamical systems, J. Roy. Stat. Soc. B 57 (4) (1995) 613–658.
- [9] T.L. Lai, Information bounds and quick detection of parameter changes in stochastic systems, IEEE Trans. Inform. Theory 44 (7) (1998) 2917–2929.
- [10] T.L. Lai, J.Z. Shan, Efficient recursive algorithms for detection of abrupt changes in signals and control systems, IEEE Trans. Automat. Control 44 (5) (1999) 952–966.
- [11] G. Lorden, Procedures for reacting to a change in distribution, Ann. Math. Stat. 42 (1971) 1897–1908.
- [12] G. Lorden, Open-ended tests for Koopman-Darmois families, Ann. Stat. 1 (1973) 633–643.
- [13] I. Nikiforov, V. Varavva, V. Kireichikov, Application of statistical fault detection algorithms to navigation system monitoring, Automatica 29 (5) (1993) 1275–1290.
- [14] I.V. Nikiforov, Modification and analysis of the cumulative sum procedure, Automat. Remote Control (Avtomatika i Telemekhanika) 41 (9, part 1) (1980) 1247–1252.
- [15] I.V. Nikiforov, On a class of quadratic tests for detection of abrupt changes in signals and systems, American Control Conference Vol. 2, Philadelphia, Pennsylvania, USA, June 24–26, 1988, pp. 751–755.
- [16] I.V. Nikiforov, On first order optimality of a change detection algorithm in a vector case, Automat and Remote Control (Avtomatika i Telemekhanika) 55 (1) (1994) 68–82.
- [17] I.V. Nikiforov, A generalized change detection problem, IEEE Trans. Inform. Theory 41 (1) (1995) 171–187.

- [18] I.V. Nikiforov, New optimal approach to Global Positioning System/Differential Global Positioning System integrity monitoring, *J. Guidance Control Dyn.* 19 (5) (September–October 1996) 1023–1033.
- [19] I.V. Nikiforov, Quadratic tests for detection of abrupt changes in multivariate signals, *IEEE Trans. Signal Process.* 47 (9) (September 1999) 2534–2538.
- [20] I.V. Nikiforov, A suboptimal quadratic change detection scheme, *IEEE Trans. Inform. Theory* 46 (6) (2000) 2095–2107.
- [21] J.J. Pignatiello, G.C. Runger, Comparisons of multivariate CUSUM charts, *J. Qual. Technol.* 22 (3) (July 1990) 173–186.
- [22] M. Pollak, Optimal detection of a change in distribution, *Ann. Stat.* 13 (1985) 206–227.
- [23] M. Pollak, D. Siegmund, Approximations to the expected sample size of certain sequential tests, *Ann. Stat.* 3 (6) (1975) 1267–1282.
- [24] Y. Ritov, Decision theoretic optimality of the CUSUM procedure, *Ann. Stat.* 18 (1990) 1464–1469.
- [25] D. Siegmund, E.S. Venkatraman, Using the generalized likelihood ratio statistics for sequential detection of a change-point, *Ann. Stat.* 23 (1994) 255–271.
- [26] M.A. Sturza, Navigation system integrity monitoring using redundant measurements, *Navigation* 35 (1988–89) 483–501.
- [27] A. Wald, *Sequential Analysis*, Wiley, New York, 1947.
- [28] A.S. Willsky, H.L. Jones, A generalized likelihood ratio approach to detection and estimation of jumps in linear systems, *IEEE Trans. Automat. Control* 21 (1976) 108–112.
- [29] B. Yakir, A note on optimal detection of a change in distribution, *Ann. Stat.* 25 (5) (1997) 2117–2126.
- [30] Q. Yao, Asymptotically optimal detection of a change in a linear model, *Sequential Anal.* 12 (1993) 201–210.